SINGLE-MOLECULE TRACKING STUDIES OF METALLOREGULATORS IN LIVE BACTERIA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Won Jung

December 2018

SINGLE-MOLECULE TRACKING STUDIES OF METALLOREGULATORS IN LIVE BACTERIA

Won Jung, Ph. D.

Cornell University 2018

This dissertation covers two major topics. First topic is a single-molecule tracking (SMT) study of zinc-responsive Fur-family metalloregulator (Zur) to investigate Zur-DNA interactions and its transcriptional regulation in living cells; second topic is a 3-dimensional molecular diffusion simulation study which allows for direct connection between SMT data and diffusion theory, and thus extracting the intrinsic properties of Brownian motion.

Transcription regulator on-off binding to DNA constitutes a mechanistic paradigm in gene regulation, in which the repressors/activators bind to operator sites tightly while the corresponding non-repressors/non-activators do not. Another paradigm regards regulator unbinding from DNA to be a unimolecular process whose kinetics is independent of regulator concentration. Using single-molecule single-cell measurements, we find that the behaviors of the zinc-responsive uptake regulator Zur challenges these paradigms. Apo-Zur, a non-repressor and presumed non-DNA binder, can bind to chromosome tightly in live *E. coli* cells, likely at non-consensus sequence sites. Moreover, the unbinding from DNA of its apo-non-repressor and holo-repressor forms both show a biphasic, repressed-followed-by-facilitated kinetics with increasing

cellular protein concentrations. The facilitated unbinding likely occurs via a ternary complex formation mechanism; the repressed unbinding is *first-of-its-kind* and likely results from protein oligomerization on chromosome, in which an inter-protein salt-bridge plays a key role. This biphasic unbinding could provide functional advantages in Zur's facile switching between repression and derepression.

Single-molecule tracking (SMT) of fluorescently-tagged cytoplasmic proteins can provide valuable information on the underlying biological processes in living cells via subsequent analysis of the displacement distributions. However, the confinement effect originated from the small size of a bacterial cell skews the protein's displacement distribution and complicates the quantification of the intrinsic diffusive behaviors. Using the inverse transformation method, we convert the skewed displacement distribution (for both 2D and 3D imaging conditions) back to that in free space for systems containing one or multiple (non)interconverting Brownian diffusion states, from which we can reliably extract the number of diffusion states as well as their intrinsic diffusion coefficients and respective fractional populations. We further demonstrate a successful application to experimental SMT data of a transcription factor in living *E. coli* cells. This work allows a direct quantitative connection between cytoplasmic SMT data with diffusion theory for analyzing molecular diffusive behavior in live bacteria.

BIOGRAPHICAL SKETCH

Won Jung was born on Mar. 03, 1983 in Seoul, South Korea. He obtained his Bachelor of Science (B.S.) degree in Chemistry Department at Chung-Ang in Seoul City in 2008. He did his undergraduate thesis in theoretical chemistry under the supervision of Prof. Jaeyoung Sung. After college, he continued on his research under the supervision of Prof. Jaeyoung Sung, and obtained his Master of Science (M.S.) degree in 2010. The master thesis's title is "Novel Chemical Kinetics for a Single Enzyme Reaction: Relationship between Substrate Concentration and the Second moment of Enzyme Reaction Time". After MS, he moved to Cornell University in Ithaca, New York to obtain his doctoral degree (Ph.D.) in Chemistry and Chemical Biology under the supervision of Prof. Peng Chen. He extensively used single-molecule fluorescence imaging of protein-DNA interactions in living cells to understand metal-responsive gene regulation mechanisms in bacteria, in combination with genetic manipulations.

ACKNOWLEDGMENTS

I would like to thank my advisor, Prof. Peng Chen, for his mentorship and support throughout my graduate career. His mentoring helped broaden not only my technical skills but intellectual horizons as a scientist. He also made me confident in not only research but life.

I thank my committee members, Prof. Hening Lin and Prof. Warren Zipfel, for their valuable inputs, critical insights and discussions.

I thank my research collaborators: Ace George Santiago for effectively teaching me the basics of molecular biology techniques and providing basic cell strains and templates, Tai-Yen Chen for building the PALM microscope and providing iQ-PALM source codes, Kushal Sengupta, Lauren Genova, Chandra Joshi, Bing Fu, Mahdi Hesari, Udit Chakraborty, Łukasz Krzeminski, Feng Yang, Danya Smart for discussion. Prof. John Helmann (Cornell University) for scientific discussion.

I thank Yimon Aye (Cornell University) and Cynthia Kinsland (Cornell University) for access to biology facilities.

I thank the Department of Chemistry and Chemical Biology for the Teaching Assistantships and the National Institutes of Health for the financial support of our research.

I thank lastly but heartily, my wife Jahyon Park for mental and unconditional support and my parents, Sung-mi and Kwang-mo for selfless sacrifices for my living.

Thank you all.

TABLE OF CONTENTS

CHAPTER 1

BIPHASIC UNBINDING OF ZUR FROM DNA FOR TRANSCRIPTION (DE)REPRESSION IN LIVE BACTERIA

1.1 INTRODUCTION

Transcriptional regulation in cells is generally orchestrated by regulators, which, upon binding to operator sites, either block the binding of RNA polymerase (RNAP) leading to repression (i.e., repressors) or recruit RNAP leading to activation (i.e., activators)[1, 2]. One mechanistic paradigm for these regulators is an on-off model in which they bind to their cognate operator sites tightly, while their corresponding non-repressor/non-activator forms have insignificant affinity to DNA and stay predominantly in the cytoplasm. Some exceptions recently emerged. For example, IscR, a member of the MarA/SoxS/Rob family of transcription regulators in *E. coli*, is a repressor in its holo-form (i.e., containing a Fe-S cluster); its apo-form, generally thought to not bind DNA, was shown to bind DNA motifs different from its holo-repressor form[3, 4].

Derepression or deactivation subsequently comes from the unbinding of the regulator from the operator site. Here another mechanistic paradigm exists regarding the kinetics of regulator unbinding, which is presumed to be a unimolecular reaction (i.e., spontaneous unbinding), whose first-order rate constant is independent of surrounding regulator concentration. However, recent *in vitro* single-molecule and bulk measurements uncovered facilitated unbinding, in which the first-order unbinding rate constant increases with increasing protein concentrations[5]. These proteins include nucleoid associated proteins that bind double-stranded DNA nonspecifically[6], replication protein A that binds single-stranded DNA

nonspecifically[7], and DNA polymerases[8, 9]. We also discovered that CueR and ZntR, two MerR-family metal-sensing transcription regulators that bind to their cognate promoter sequences specifically, also show facilitated unbinding[10]. Using single-molecule tracking (SMT) and single cell quantification of protein concentration (SCQPC) that connect protein-DNA interaction kinetics with cellular protein concentrations quantitatively, we further showed that the facilitated unbinding of CueR and ZntR also operate in living *E. coli* cells[11]. A mechanistic consensus emerged, involving multivalent contacts between the protein and DNA[5], which enables the formation of ternary complexes as intermediates that subsequently give rise to concentration-enhanced protein unbinding kinetics.

Here we report a SMT and SCQPC study of Zur, a Fur-family homodimeric zinc-uptake regulator, whose Zn^{2+}-bound holo-form binds to its cognate operator site with nM affinity and represses the transcription of zinc uptake genes under zinc stress[12-15]; its apo-form is a non-repressor. We found that in living *E. coli* cells, Zur's interactions with DNA challenge the above two paradigms. First, apo-Zur, long thought to not bind DNA, can bind to chromosome tightly, likely at non-consensus sites. Second and more strikingly, the unbinding of both apo- and holo-Zur from chromosome not only show facilitated unbinding with increasing cellular protein concentrations, but also exhibit repressed unbinding at lower concentrations, giving a first-of-its-kind biphasic unbinding behavior. The repressed unbinding of Zur likely stems from Zur oligomerization on DNA, where an inter-dimer salt bridge plays a key role, and it likely facilitates transcription switching between repression and depression in cells.

1.2 RESULTS

1.2.1 SMT and SCQPC identify a tight DNA-binding state for both holo- and apo-Zur in cells

To visualize individual Zur proteins in *E. coli* cells, we fused the photoconvertible fluorescent protein mEos3.2[16, 17] to its C-terminus creating Zur^{mE}, either at its chromosomal locus to have physiological expression or in an inducible plasmid in a *zur* deletion strain to have a wider range of cellular protein concentrations (Methods). This Zur^{mE} fusion-protein is intact and as functional a repressor as the wild-type (WT) in the cell under Zn stress growth conditions (Supplementary Figure 1a-b).

Using sparse photoconversion and time-lapse stroboscopic imaging, we tracked the motions of photoconverted Zur^{mE} proteins individually in single *E. coli* cells at tens of nanometer precision until their mEos3.2 tags photobleached (Figure 1a). This SMT allows for measuring Zur^{mE}'s mobility, which reports on whether the molecule is freely diffusing in the cell or bound to DNA. We repeated this photoconversion and SMT cycle 500 times for each cell, during which we counted the number of tracked protein molecules. We then used the SCQPC protocol to quantify the remaining number of Zur^{mE} protein molecules in the same cell[11], eventually determining the Zur^{mE} concentration in each cell (i.e., $[Zur^{mE}]_{cell}$). This single-cell protein quantitation allowed for sorting the cells into groups of similar protein concentrations and subsequently examining protein-concentration–dependent processes, without being limited by the large cell-to-cell heterogeneity in protein expression.

We first examined Zur^{mE}_{apo} whose regulatory Zn-binding site was mutated (i.e., C88S) to make it permanent apo and a non-repressor[15] (Supplementary Figure 1b). To quantify its

mobility in cells, we determined the distribution of its displacement length *r* between successive images and the corresponding cumulative distribution function (CDF) of *r* for each cell group having similar cellular Zur^{mE}_{apo} concentrations (Figure 1b-c). Global analysis of these CDFs across all cellular protein concentrations resolved minimally three Brownian diffusion states with *effective* diffusion constants of ~5.0 ± 0.5, 0.82 ± 0.05, and 0.040 ± 0.003 $\mu m^2\ s^{-1}$ (Figure 1b-c; Methods). No subcellular localization or protein aggregation was observed; therefore, these two aspects are not the reasons for the presence of these three diffusion states. On the basis of their diffusion constants and previous studies of transcription regulator diffusion in *E. coli* cells[11, 18-21], we assigned the fastest diffusion state as Zur^{mE}_{apo} proteins freely diffusing (FD) in the cytoplasm, the medium diffusion state as those nonspecifically bound (NB) to and moving on chromosome, and the slowest state as those tightly bound (TB) to the chromosome, whose small effective diffusion constant (~0.040 $\mu m^2\ s^{-1}$) reflects chromosome dynamics[19, 22] and measurement uncertainties. Control measurements on the free mEos3.2 further support the assignment of the FD state, as we reported[11].

The resolution of CDFs of *r* also gave the fractional populations of the three states across the range of cellular protein concentrations (Figure 1d). With increasing $[Zur^{mE}_{apo}]_{cell}$, the fractional population of the FD state increases, while that of the TB state decreases. These trends further support their assignments because, with increasing cellular protein concentrations, more proteins compete for the limited number of tight binding sites on chromosome, leading to smaller fractional populations of the TB state and larger fractions of the FD state.

The presence of a significant fraction of the tight DNA-binding state, even at low

cellular protein concentrations, is surprising for Zur^{mE}_{apo} (e.g., ~32% at $[Zur^{mE}_{apo}]_{cell}$ ~ 60 nM; 1 nM in an *E. coli* cell corresponds to ~1 protein copy), as apo-Zur is a non-repressor. Furthermore, previous gel shift assay showed that *E. coli* apo-Zur does not bind to operator sites (i.e., K_D > 300 nM at the *znuABC* promoter)[15], and for *B. subtilis*, its apo-Zur's binding affinity to operator sites is ~1000 times weaker than its holo-form[23]. We hypothesized that the TB state of Zur^{mE}_{apo} likely comes from its binding to non-operator sites (i.e., non-consensus sequence sites; see later).

We next examined Zur^{mE} in cells stressed with 20 μM Zn^{2+} in the medium. This Zn^{2+} concentration can evoke maximal repression of *zur* regulons (Supplementary Note 2.3). Therefore, most of Zur proteins in the cell should be metallated, mimicking the holo repressor form (i.e., Zur^{mE}_{Zn}). The same three diffusion states are resolved in the CDFs of *r* across all cellular protein concentrations (Supplementary Note 4.2). In contrast to the case for Zur^{mE}_{apo}, the TB state of Zur^{mE}_{Zn} is expected here because holo-Zur binds specifically to consensus operator sites within Zur-regulated promoters. Expectedly, the fractional population of the FD state of Zur^{mE}_{Zn} increases with increasing $[Zur^{mE}_{Zn}]_{cell}$, whereas that of the TB state decreases (Figure 1d).

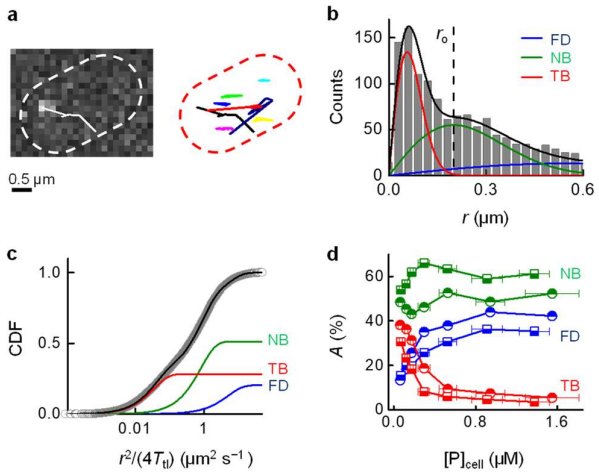


Figure 1. SMT of Zur in living cells. a, Left: exemplary fluorescence image of a single Zur^{mE}_{apo} protein in a live *E. coli* cell overlaid with its position trajectory (solid line). Right: overlay of many trajectories. Dash lines: cell boundary. **b**, Histogram of displacement length r per time-lapse (40 ms) of > 1,400 tracked Zur^{mE}_{apo} proteins at 124 ± 15 nM. Solid lines: the overall fitted distribution (black), and the resolved FD (blue), NB (green), and TB (red) diffusion states (Supplementary Note 4). Vertical dashed line: r_o = 0.2 μm for extracting residence times as in Figure 2a. **c**, Cumulative-distribution-function (CDF) of r (plotted against $\frac{r^2}{4T_{tl}}$) as in **b**. Lines: overall fit (Eq. (3)) and three resolved diffusion states with effective diffusion constants (and fractional populations): D_{FD} = 5.0 ± 0.5 μm^2 s^{-1} (21.7 ± 0.4%), D_{NB} = 0.8 ± 0.05 μm^2 s^{-1} (48.8 ± 0.4%), and D_{TB} = 0.040 ± 0.003 μm^2 s^{-1} (30.1 ± 0.5%). **d**, Fractional populations of FD, NB, and TB states for Zur^{mE}_{apo} (half-solid squares) and Zur^{mE}_{Zn} (half-solid circles) vs. the cellular protein concentrations.

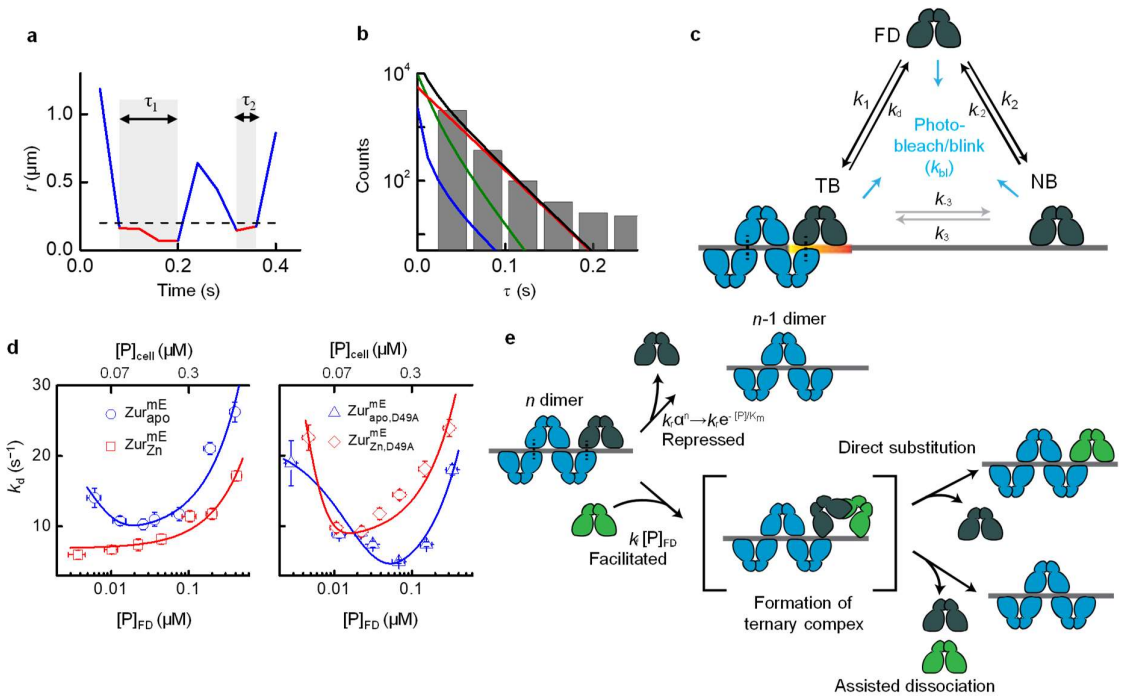


Figure 2. Biphasic unbinding kinetics of Zur from TB sites on chromosome. a, Time trajectory of displacement length r per time-lapse from a single Zur^{mE}_{apo} protein. Two microscopic residence time τ shown in gray shades; dashed horizontal line: displacement threshold $r_o = 0.2$ μm (vertical dashed line in Figure 1b). **b**, Histogram of τ for Zur^{mE}_{apo} at the cellular concentration of 124 ± 15 nM. Black line: fitting with Eq. (4). Contributions of the three diffusion states are plotted, as color-coded in Figure 1b-c. **c**, Three-state model of a single Zur protein interacting with DNA in a cell. k's are the rate constants. **d**, Protein-concentration-dependent k_d for Zur^{mE}_{apo} and Zur^{mE}_{Zn} (left) and their corresponding D49A salt-bridge mutants (right). Bottom/top axis refers to free/cellular protein concentration, respectively. Lines are fits with Eq. (2). All error bars are s.d. **e**, Schematic molecular mechanisms for biphasic unbinding of Zur from a TB site. A bound Zur protein (dark blue) within an oligomer on DNA can unbind following either a repressed pathway (top) due to the presence of (n-1) proteins nearby or a

facilitated pathway (bottom) upon binding another protein (green) to form an intermediate ternary complex, which then proceeds through direct substitution or assist dissociation pathway. Black dashed lines denote salt-bridge interactions.

1.2.2 Concentration-dependent biphasic unbinding kinetics of Zur from DNA

To probe Zur–DNA interaction dynamics, we examined the r versus time t trajectories of individual Zur proteins inside cells. These trajectories show clear transitions between large and small r values (Figure 2a): the small r values are expected to be dominated by instances of Zur tightly bound to chromosome (i.e., TB state). We set an upper threshold r_0 (= 0.2 μm), below which >99.5% of the TB states are included based on the resolved distributions of r (Figure 1b), to select these small displacements and obtain estimates of the individual residence time τ of a single Zur protein at a chromosomal tight binding site (Figure 2a). Each τ starts when r drops below r_0 and ends when r jumps above r_0 (e.g., τ's in Figure 2a), which are expected to reflect dominantly protein unbinding from DNA, or when the mEos3.2-tag photobleaches/blinks.

We analyzed trajectories from many cells of similar cellular Zur concentrations to obtain their corresponding distribution of τ (Figure 2b). We used a quantitative three-state model (i.e., FD, NB, and TB states; Figure 2c) to analyze the distribution of τ, in which the contributions of FD and NB states are deconvoluted (Eq. (4); approximations and validations of this model in Supplementary Note 5)[11]. This model also accounts for mE photobleaching/blinking kinetics, determined from the fluorescence on-time distribution of

SMT trajectories (Supplementary Figure 8). This analysis gave k_d, the apparent first-order unbinding rate constant of Zur from a tight binding site on the chromosome, for each group of cells having similar cellular Zur concentrations.

Strikingly, k_d for Zur^{mE}_{apo} shows a biphasic, repressed-followed-by-facilitated behavior: it initially decreases with increasing free (or total) cellular Zur concentration (i.e., repressed), reaching a minimum at ~130 nM; it then increases toward higher protein concentrations (i.e., facilitated; Figure 2d, left, blue points). This biphasic behavior is also apparent in the simple averages of residence time τ or by analyzing the distributions of τ that merely takes into account mE photobleaching/blinking (Supplementary Figure 9a). The facilitated unbinding of Zur^{mE}_{apo} is analogous to those of CueR and ZntR, two MerR-family metalloregulators that we discovered *in vitro* and in living cells[10, 11]; the repressed unbinding of Zur^{mE}_{apo} is a *first-of-its-kind* discovery, however.

In contrast, k_d for Zur^{mE}_{Zn} only shows the facilitated unbinding within the accessible cellular protein concentration range (~30 to ~900 nM) — it increases consistently with increasing cellular protein concentrations (Figure 2d, left, red points). The different behaviors of Zur^{mE}_{Zn} from that of Zur^{mE}_{apo} indicate that we could indeed observe the behaviors of the holo-repressor.

1.2.3 Mechanism of biphasic unbinding of Zur from DNA

Amid the biphasic unbinding of Zur from DNA (Figure 2d, left), the concentration-facilitated unbinding at higher protein concentrations is analogous to those of CueR and ZntR[11]. There it stems from an assisted dissociation pathway, in which an incoming protein from

solution helps an incumbent protein on DNA to unbind, or a direct substitution pathway, in which the incoming protein directly replaces the incumbent one (Figure 2e, lower)[10, 11]. The rates of both pathways depend linearly on the free protein concentration, and both likely occur through a common ternary protein$_2$–DNA complex, in which the two homodimeric proteins each use one DNA-binding domain to bind to half of the dyad recognition sequence[5, 24]. As Zur is also a homodimer, Zur also could form this ternary complex and undergo assisted dissociation or direct substitution, leading to its concentration-facilitated unbinding from DNA.

Regarding the repressed unbinding of apo-Zur in the lower concentration regime, we propose that it likely results from protein oligomerization around the DNA binding site, in which the number of proteins in the oligomer increases with increasing protein concentration and the resulting protein-protein interactions contribute to additional stabilization, thereby repressing protein unbinding rate (Figure 2e, upper). (The facilitated unbinding later takes over when the protein concentration reaches a high enough level.) Two evidences support our oligomerization proposal: (1) Crystallography showed that two *E. coli* Zur dimers can bind to a short cognate DNA sequence[15]. (2) DNA footprinting showed that *S. coelicoror* Zur forms oligomers around its recognition sites, containing greater than 4 dimers[25].

To further support this oligomerization proposal, we examined the spatial distribution in the cell of Zur's residence sites at its TB state; these residence sites correspond to the r_0-thresholded small displacements (Figure 2a; Supplementary Note 8). For comparison, we further simulated an equal number of sites randomly distributed in a cell of the same size (Supplementary Note 8.1). We then examined their pair-wise distance distributions (PWD), in

which Zur oligomerization at chromosomal binding sites should lead to more populations at shorter pair-wise distances. This PWD for Zur^{mE}_{apo} indeed shows a higher population at distances shorter than ~500 nm relative to the simulated random sites (Figure 3a). However, at the distance scale of a few hundred nanometers, the compaction of chromosome also contributes to the PWD of residence sites[11]. To decouple the contribution of protein oligomerization from chromosome compaction, we examined the fraction of residence sites within a radius threshold *R*. At small *R* (e.g., <100 nm), the contribution of Zur oligomerization to this fraction should dominate over chromosome compaction, as oligomerization is at molecular scale whereas the most compact chromosome in a *E. coli* cell is still around hundreds of nanometer in dimension[11, 26]. At any specified *R* (e.g., 200 nm), the fraction of Zur^{mE}_{apo} residence sites within the radius *R* increases expectedly with increasing cellular protein concentrations (Figure 3b, red points), because higher protein concentrations gave higher sampling frequency of residence sites. More important, at lower *R* (e.g., 100 nm), the fraction of Zur^{mE}_{apo} residence sites is larger than that of simulated random sites (Fig 3b, red vs. blue points), and their ratio is larger at lower protein concentrations (Figure 3b, green points). The average ratio of the fraction of Zur^{mE}_{apo} residence sites over that of the simulated random sites is always greater than 1, and it becomes larger at smaller *R* down to <70 nm (Figure 3c; note our molecular localization precision is ~20 nm; Supplementary Note 3), supporting Zur^{mE}_{apo} oligomerization at chromosomal tight binding sites at the nanometer scale.

We formulated a quantitative kinetic model to describe the biphasic unbinding of Zur^{mE}_{apo}. It considers both oligomerization at a TB site and facilitated unbinding via a ternary $protein_2$-DNA complex (Figure 2c and e; Supplementary Note 6). The microscopic unbinding rate

constant $k_d^{(n)}$ from a TB site with n Zur_{apo}^{mE} dimers bound as an oligomer comprises three terms:

$$k_d^{(n)} = k_o + k_r \alpha^n + k_f [P]_{FD} \quad \textbf{(1)}$$

k_o is a first-order intrinsic unbinding rate constant. The $k_r\alpha^n$ term accounts for the repressed unbinding from protein oligomerization, where a first-order rate constant k_r is attenuated by the factor α ($0 < \alpha < 1$) to the exponent of n, which depends on the cellular protein concentration and has a maximal value of n_0, the oligomerization number. The third term describes the facilitated unbinding, with k_f being a second-order rate constant and $[P]_{FD}$ being the concentration of freely diffusing Zur dimers in the cell, as reported for CueR/ ZntR[11]. In the limit of weak oligomerization and low free protein concentrations, the apparent unbinding rate constant k_d from any TB site is:

$$k_d = \left\langle k_d^{(n)} \right\rangle = k_o^{off} + k_r \left(e^{-[P]_{FD}/K_m} - 1 \right) + k_f [P]_{FD} \quad \textbf{(2)}$$

$K_m = \frac{k_o^{off}}{k_1(1-\alpha)}$; it has the units of protein concentration, reflecting the effective dissociation constant of the protein oligomer on the chromosome. $k_o^{off} = k_o + k_r$; it is a first-order spontaneous unbinding rate constant at the limit of zero cellular protein concentration. Equation (2) satisfactorily fits the biphasic unbinding kinetics of Zur_{apo}^{mE} (Figure 2d, left), giving the associated kinetic parameters (Table 1 and Supplementary Table 6). In particular, K_m of Zur_{apo}^{mE} is ~5 nM, indicating that apo-Zur can oligomerize on chromosome at its physiological concentrations in the cells (Figure 4a).

The same model also allowed for analyzing the relative populations of FD, NB, and TB states of Zur across all cellular protein concentrations, giving additional thermodynamic

and kinetic parameters (Table 1, and Supplementary Table 6). Strikingly, the dissociation constant K_{d1} of Zur^{mE}_{apo} at TB sites of DNA is ~11 nM, merely ~2 times weaker than that of Zur^{mE}_{Zn} (K_{d1} ~5 nM). This is *not* expected because apo-Zur, in both *E. coli* and *B. subtilis*, was shown to have no significant affinity to the consensus sites recognized by holo-Zur[15, 23]. Therefore, the high affinity of Zur^{mE}_{apo} at the TB state suggests that inside cells, apo-Zur likely bind tightly to other, non-consensus sites in the chromosome. This likelihood is supported by a ChIP-seq analysis in *B. subtilis*, which showed Zur can bind tightly to many locations in the chromosome that do not share consensus with the known recognition sites (although it was undefined whether the detected bindings there were by apo- or holo-Zur)[27].

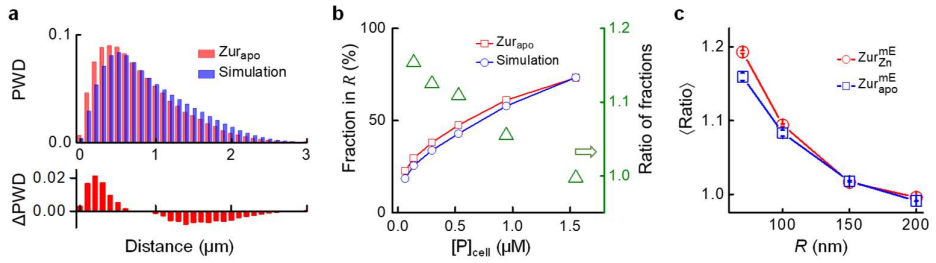


Figure 3. Spatial analysis of Zur's residence sites. a, Normalized pair-wise distance distributions (PWD) of residence sites for Zur^{mE}_{apo} and for simulated random sites in the cell (top), and the difference of Zur^{mE}_{apo} from simulation (bottom). **b**, Fraction of residence sites within a radius threshold R (= 100 nm, left axis) for Zur^{mE}_{apo} and for simulated random sites as a function of cellular protein concentration. Their ratio (Zur^{mE}_{apo} vs. simulation) is plotted against the right axis. **c**, Dependence of the average ratio in **b** across all protein concentrations as a function of the radius threshold R for Zur^{mE}_{apo} and Zur^{mE}_{Zn}.

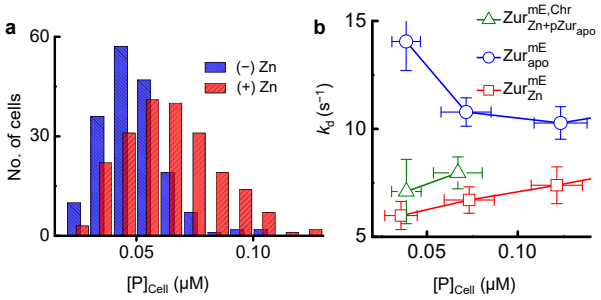


Figure 4. Zur behaviors within the physiological range of cellular protein concentrations. a, Distribution of the chromosomally expressed Zur^{mE} concentration in the cell with (+) and without (−) Zn stress in the medium. **b**, Dependence of k_d on the protein concentration in the cell for Zur^{mE}_{apo}, Zur^{mE}_{Zn}, and for Zur^{mE}_{Zn} together with a plasmid expressing Zur_{apo} (i.e. $Zur^{mE,Chr}_{Zn+PZur_{apo}}$) when the mE-tagged Zur is only encoded on the chromosome. The blue circles and red squares for Zur^{mE}_{apo} and Zur^{mE}_{Zn} are part of data in Figure 2d (left).

1.2.4 Molecular basis of repressed unbinding

Our model of Zur oligomerization at TB sites was based partly on the structure of two holo-Zur dimers bound to a cognate DNA, which showed two inter-dimer D49−R52 salt bridges[15]. To probe the role of these salt bridges in Zur oligomerization, we made the D49A mutation, known to disrupt the interactions[15]. For apo-Zur, the resulting mutant $Zur^{mE}_{apo,\ D49A}$ still exhibits the biphasic unbinding behavior, however the minimum of the apparent unbinding rate constant $\underline{k}_d$ shifted to a higher cellular protein concentration (Figure 2d, right). Its K_m is 16.2 ± 7.5 nM, three times larger than that of Zur^{mE}_{apo} (Table 1), indicating a weakened oligomerization affinity and thus a significant role of these salt bridges.

More strikingly, for Zur^{mE}_{Zn}, which only showed facilitated unbinding (Figure 2d, left), the resulting mutant $Zur^{mE}_{Zn,\ D49A}$ clearly shows biphasic unbinding with K_m = 3.2 ± 1.9 nM (Figure 2d, right; Table 1). Therefore, holo-Zur also possesses repressed unbinding kinetics — it was invisible for Zur^{mE}_{Zn} likely because its K_m is smaller than the low limit of accessible cellular protein concentrations (~3 nM), but emerges after the D49A mutation, which further supports the importance of the salt bridges in Zur oligomerization and repressed unbinding behaviors.

1.3 DISCUSSION

We have uncovered that the Fur-family Zn^{2+}-sensing transcription regulator Zur exhibits two unusual behaviors that challenge conventional paradigms of regulator-chromosome interactions. First, apo-Zur, the non-repressor form and a long-presumed non-DNA binder, can actually bind to chromosome tightly, likely at different locations from the consensus sequence recognized by holo-Zur, the repressor form. This tight chromosome binding by apo-Zur challenges the paradigm of regulator on-off model for transcription repression (or activation)[1, 2]. Second, the unbinding kinetics of both apo- and holo-Zur not only exhibit facilitated unbinding, a newly discovered phenomenon for a few DNA-binding proteins[6, 7, 9, 28], but also show repressed unbinding, a *first-of-its-kind* phenomenon that likely results from Zur oligomerization on chromosome, facilitated by inter-dimer salt bridges. Overall, Zur has biphasic unbinding kinetics from chromosome with increasing cellular protein concentrations, which challenges the paradigm of protein unbinding being typically unimolecular processes whose first-order rate constants do not depend on the protein

concentration.

To probe whether the biphasic unbinding of Zur occurs within the physiological cellular protein concentrations, we quantified cellular Zur^{mE} concentration when it is encoded only at the chromosomal locus (Figure 4a). In minimal medium without Zn stress, the cellular Zur^{mE}, which is mostly in the apo-form, ranges from ~24 to 108 nM (mean = 50 ± 14 nM), within which apo-Zur unbinding from TB sites is in the repressed unbinding regime and slows down by ~42% from the lowest to the highest protein concentration (Figure 4b). When stressed by 20 μM Zn^{2+}, the cellular Zur^{mE}, now mostly in the holo-form, ranges from ~26 to 124 nM (mean = 63 ± 20 nM), reflecting an average of ~28% protein concentration increase induced by Zn stress. In this protein concentration range, holo-Zur is already in the facilitated unbinding regime, and its unbinding rate from a recognition site can increase by ~36% (Figure 4b).

Within the physiological protein concentration range, the opposite dependences of unbinding kinetics on the cellular protein concentration between apo- and holo-Zur could provide functional advantages for an *E. coli* cell to repress or de-repress Zn uptake genes. When cell encounters environmental Zn stress that demands strong repression of Zn uptake, the cellular concentration of Zur swings upward and it becomes dominantly in the holo-repressor form. The unbinding of holo-repressor from recognition sites could be facilitated by its increasing concentration (Figure 5a), but the facilitated unbinding via direct substitution by another holo-repressor has no functional consequence while facilitated unbinding via assisted dissociation will be immediately compensated by a rebinding of a holo-repressor (the rebinding would occur within ~0.014 s; Supplementary Note 7). For those cellular Zur in the apo non-repressor form, its unbinding from DNA slows down, keeping them longer (i.e., stored) at non-

consensus chromosomal sites (Figure 5b). On the other hand, when cell transitions to a Zn-deficient environment that demands derepression of Zn uptake, the cellular Zur protein concentration goes down. Here unbinding of the holo-repressor would be slower (Figure 5c), which is undesirable for derepression, while the unbinding of the apo-form would become faster, releasing them from the non-consensus "storage" sites on the chromosome into the cytosol (Figure 5d). If the cytosolic apo-Zur could possibly facilitate the unbinding of holo-Zur from promoter recognition sites (e.g., through assisted dissociation), it would give a more facile transition to derepression. To support this possibility, we measured the apparent unbinding rate constant k_d for chromosomally encoded Zur^{mE}_{Zn} in cells that contains a plasmid encoding an untagged Zur_{apo} mutant (i.e., C88S). When the expression of this Zur_{apo} mutant is induced, k_d of Zur^{mE}_{Zn} increases by ~28% at any cellular Zur^{mE}_{Zn} concentration (Figure 4b, green vs. red points), indicating that apo-Zur can indeed facilitate the unbinding of holo-Zur from recognition sites (Figure 5e).

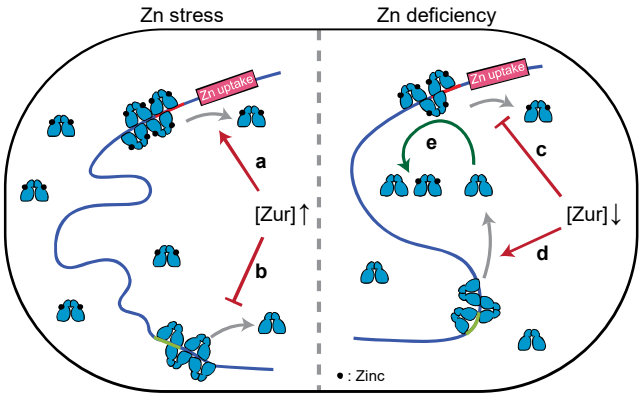


Figure 5. Functional model of holo- and apo-Zur unbinding behaviors in *E.coli* upon encountering zinc stress or deficiency. Upon zinc stress, unbinding of holo-Zur from operator site is facilitated (**a**) while that of apo-Zur from storage site is repressed (**b**) due to

increase in cellular protein concentration. Upon zinc deficiency, the facilitated unbinding of holo-Zur is attenuated (**c**) while the unbinding of apo-Zur is less repressed (**d**) due to decrease in cellular protein concentration. Released apo-Zur into cytosol could facilitate holo-Zur to unbind (**e**), helping transition to de-repression of zinc uptake.

Table 1 | Kinetic and thermodynamic parameters for Zur-DNA interaction in *E.coli* cells

	Zur^{mE}	Zur^{mE}_{apo}	Zur^{mE}_{Zn}	$Zur^{mE}_{apo,\ D49A}$	$Zur^{mE}_{Zn,\ D49A}$
k_1(nM^{-1} s^{-1}) [a]	1.90 ± 0.17	1.84 ± 0.20	1.10 ± 0.18	1.61 ± 0.58	1.30 ± 0.19
k_o^{off} (s^{-1})	25 ± 12	22 ± 21	5.4 ± 0.6	22.1 ± 1.5	36 ± 41
k_r (s^{-1})	16 ± 11	12 ± 20	n/o [b]	20.8 ± 1.3	27 ± 40
k_f(nM^{-1} s^{-1})	0.028 ± 0.005	0.044 ± 0.007	0.026 ± 0.033	0.049 ± 0.014	0.062 ± 0.010
K_m (nM)	6.0 ± 4.0	4.9 ± 7.3	n/o [b]	16.2 ± 7.5	3.2 ± 1.9
K_{d1} ($=k_o^{off}/k_1$) (nM) [a]	12.9 ± 6.2	11.7 ± 11.2	4.9 ± 1.2	13.7 ± 5.0	28 ± 20
K_{d2} ($=k_{-2}/k_2$) (nM) [a]	417 ± 35	348 ± 84	534 ± 148	209 ± 69	532 ± 134
K_{d3} ($=k_{-3}/k_3$) [a]	0.011 ± 0.002	0.023 ± 0.007	0.022 ± 0.023	0.032 ± 0.062	0.008 ± 0.006
$[D_0]_{NB}$ (nM) [a]	1144 ± 84	961 ± 205	1201 ± 287	858 ± 230	1538 ± 353
$[D_0]_{TB} \cdot n_o$ (nM) [a]	42.56 ± 0.94	34.3 ± 3.2	54 ± 14	31.6 ± 5.1	38.8 ± 3.8

[a] $n_o = 5$ was used in fitting.
[b] Not observed

Multivalent contacts with DNA, which underlie the facilitated unbinding, and salt-bridge interactions between proteins, which underlie Zur oligomerization and its repressed unbinding, are both common for protein-DNA and protein-protein interactions, respectively[5, 7, 10, 28-36]. Therefore, the biphasic unbinding behavior from DNA discovered here for Zur could be broadly relevant to many other proteins in gene regulation.

1.4 METHODS

1.4.1 Bacterial strains and sample preparation

All strains were derived from the *E.coli* BW25113 strain as detailed in Supplementary Note 1. Zur^{mE} was either encoded at its chromosomal locus via lambda-red homologous recombination[37] or in a pBAD24 plasmid in a Δ*zur* deletion strain[38]. Mutant forms of Zur (Zur^{mE}_{apo}, Zur^{mE}_{D49A}, or $Zur^{mE}_{apo,\ D49A}$) were generated via site-directed mutagenesis in pBAD24, which was introduced into the Δ*zur* strain.

All cell imaging experiments were done at room temperature in M9 medium supplemented with amino acids, vitamins, and 0.4% glycerol. 20 μM $ZnSO_4$ was used for Zn stress conditions. The cells were immobilized on an agarose pad in a sample chamber. Details in Supplementary Note 3.

1.4.2 SMT and SCQPC

SMT and SCQPC were performed on an inverted fluorescence microscope, as reported[11] (Supplementary Note 3). For SMT, inclined epi-illuminated 405 nm and 561 nm lasers photoconverted and excited single mEos3.2 molecules, respectively. 561 nm excitation-imaging were in stroboscopic mode, with 4 ms laser excitation pulses separated by 40 ms time lapse, synchronized with the camera exposure, so that the mobile proteins still appear as diffraction-limited spots. A custom-written MATLAB software was used to identify diffraction-limited fluorescence spots and fit them with two-dimensional Gaussian functions, giving ~20 nm localization precision[11, 39]. Time trajectories of positions and displacement length *r* between

adjacent images were then extracted.

SCQPC was performed after SMT. The remaining proteins were firstly photoconverted to the red form by a long 405 nm laser illumination. The total cell red fluorescence was then imaged by the 561 nm laser to determine the protein copy number, provided the average fluorescence of a single mEos3.2 from the earlier SMT. The photoconversion efficiency of mEos3.2[40] and dimeric state of Zur were accounted for. Cell volumes were determined by fitting their optical transmission image contours with the model geometry of a cylinder with two hemispherical caps.

1.4.3 Resolution of diffusion states

The effective diffusion constants and the fractional populations of diffusion states were extracted by analyzing the CDF of displacement length r per time-lapse (T_{tl} = 40 ms), using a linear combination of three diffusion terms of CDF, as reported[11] (Equation (3)). Each term is from a 2-D Brownian diffusion model[18, 41, 42], which was regularly used to analyze SMT results of proteins in bacterial and mammalian cells[18, 21, 42-46] (model justification in Supplementary Note 4).

$$\begin{aligned}\mathrm{CDF}(r) = A_{\mathrm{FD}}\left(1-\exp\left(-\frac{r^2}{4D_{\mathrm{FD}}T_{\mathrm{tl}}}\right)\right) + A_{\mathrm{NB}}\left(1-\exp\left(-\frac{r^2}{4D_{\mathrm{NB}}T_{\mathrm{tl}}}\right)\right) \\ + (1-A_{\mathrm{FD}}-A_{\mathrm{NB}})\left(1-\exp\left(-\frac{r^2}{4D_{\mathrm{TB}}T_{\mathrm{tl}}}\right)\right)\end{aligned} \quad (3)$$

We globally fitted the CDFs across groups of cells of different cellular protein concentrations, in which the diffusion constants (D's) of respective diffusion states were shared but their factional populations (A's) were allowed to vary. Three terms were always the minimal

number of diffusion states to satisfactorily fit the CDF (details in Supplementary Note 4 and Supplementary Tables 4-5).

Note these diffusion constant values are not the intrinsic ones, as they are influenced by the cell confinement effect[47], which decreases the magnitude of the apparent diffusion constant, and by the time-lapse effect of imaging, where longer time lapse gives apparently smaller diffusion constants; both of these effects are most significant on the FD state, less on the NB state, and negligible on the TB state, and were evaluated quantitatively in a previous study of metal-responsive transcription regulators of a different family[11].

1.4.4 Determination and analysis of k_d

A three-state (FD, NB, and TB state) kinetic model, including the interconversion between states and photobleaching/blinking rates (Figure 2c), was used to analyze the distribution of residence times (upper thresholded by r_0; Figure 2a) at chromosomal TB sites to extract the apparent unbinding rate constant k_d. The respective residence time distribution functions $\varphi(\tau)$ for the FD, NB, and TB states with given diffusion constants (D's), the unbinding rate constant from the NB state k_{-2}, and photobleaching/blinking rate constant k_{bl} were derived to fit the τ distribution with the overall distribution function $\varphi_{all}(\tau)$ (Eq. (4); Supplementary Note 5).

$$\varphi_{all}(\tau) = A_{FD}\varphi_{FD}(\tau)+A_{NB}\varphi_{NB}(\tau)+A_{TB}\varphi_{TB}(\tau) \quad (4)$$

$$\varphi_{FD}(\tau)=\left[\frac{r_o^2}{4D_{FD}\tau^2}\exp\left(-\frac{r_o^2}{4D_{FD}\tau}\right)+k_{eff}^{FD}\left(1-\exp\left(-\frac{r_o^2}{4D_{FD}\tau}\right)\right)\right]\exp\left(-k_{eff}^{FD}\tau\right) \quad (5)$$

$$\varphi_{\mathrm{NB}}(\tau)=\left[\frac{r_{\mathrm{o}}^2}{4D_{\mathrm{NB}}\tau^2}\exp\left(-\frac{r_{\mathrm{o}}^2}{4D_{\mathrm{NB}}\tau}\right)+k_{\mathrm{eff}}^{\mathrm{NB}}\left(1-\exp\left(-\frac{r_{\mathrm{o}}^2}{4D_{\mathrm{NB}}\tau}\right)\right)\right]\exp\left(-k_{\mathrm{eff}}^{\mathrm{NB}}\tau\right) \quad (6)$$

$$\varphi_{\mathrm{TB}}(\tau)=k_{\mathrm{eff}}^{\mathrm{TB}}\exp\left(-k_{\mathrm{eff}}^{\mathrm{TB}}\tau\right) \quad (7)$$

Here $k_{\mathrm{eff}}^{\mathrm{FD}}=k_{\mathrm{bl}}\frac{T_{\mathrm{int}}}{T_{\mathrm{tl}}}$, $k_{\mathrm{eff}}^{\mathrm{ND}}=k_{\mathrm{bl}}\frac{T_{\mathrm{int}}}{T_{\mathrm{tl}}}+k_{\text{-}2}$, $k_{\mathrm{eff}}^{\mathrm{TB}}=k_{\mathrm{bl}}\frac{T_{\mathrm{int}}}{T_{\mathrm{tl}}}+k_{\mathrm{d}}$, and A_i is the fractional population of i^{th}-state.

The dependence of k_{d} on the cellular free diffusing protein concentration $[\mathrm{P}]_{\mathrm{FD}}$ was analyzed with Eq. (2), containing three terms representing spontaneous, repressed, and facilitated unbinding with the corresponding rate constants $k_{\mathrm{o}}^{\mathrm{off}}$, k_{r}, and k_{f}, respectively (derivation in Supplementary Note 6).

1.4.5 Analysis of relative populations

The same three-state kinetic model (Figure 2c) was used to analyze the relative populations of FD, NB, and TB states of Zur across all cellular protein concentrations. Oligomerization/deoligomerization of Zur at a TB site was modeled as 1-D sequential binding/unbinding, analogous to the Brunauer-Emmett-Teller multilayer-adsorption theory[48] but with a limited number n_0 of binding site and merely one binding rate constant k_1 (see Supplementary Note 7 for detailed derivation). Quasi-equilibrium approximation of interconversion among states was used, which approximates that the timescale of interconversion between states (~ms) are much shorter than the experimental imaging time (~hours). The kinetic parameters are then related to the relative concentrations of the proteins at three diffusion states.

$$\frac{[\mathrm{PD}]_{\mathrm{TB}}}{[\mathrm{P}]_{\mathrm{FD}}} = \frac{k_1[\mathrm{D}_0]_{\mathrm{TB}}}{k_{\mathrm{d}}} \frac{\partial \ln F_{\mathrm{TB}\leftarrow\mathrm{FD}}\left(x_{\mathrm{TB}\leftarrow\mathrm{FD}}\right)}{\partial x_{\mathrm{TB}\leftarrow\mathrm{FD}}} \quad \textbf{(8)}$$

$$\frac{[\mathrm{PD}]_{\mathrm{TB}}}{[\mathrm{P}]_{\mathrm{NB}}} = \frac{k_3[\mathrm{D}_0]_{\mathrm{TB}}}{k_{-3}\left([\mathrm{D}_0]_{\mathrm{NB}} - [\mathrm{PD}]_{\mathrm{NB}}\right)} \frac{\partial \ln F_{\mathrm{TB}\leftarrow\mathrm{NB}}\left(x_{\mathrm{TB}\leftarrow\mathrm{NB}}\right)}{\partial x_{\mathrm{TB}\leftarrow\mathrm{NB}}} \quad \textbf{(9)}$$

$$\frac{[\mathrm{PD}]_{\mathrm{NB}}}{[\mathrm{P}]_{\mathrm{FD}}} = \frac{k_2[\mathrm{D}_0]_{\mathrm{NB}}}{k_{-2} + k_2[\mathrm{P}]_{\mathrm{FD}}} \quad \textbf{(10)}$$

Here $[\mathrm{P}]_{\mathrm{FD}}$, $[\mathrm{PD}]_{\mathrm{NB}}$, and $[\mathrm{PD}]_{\mathrm{TB}}$ are the cellular protein concentrations of FD, NB, and TB states, respectively. $F_{\mathrm{TB}\leftarrow\mathrm{j}}\left(x_{\mathrm{TB}\leftarrow\mathrm{j}}\right) \equiv \sum_{i=0}^{n_0} x_{\mathrm{TB}\leftarrow\mathrm{j}}^{i}, \mathrm{j} \in [\mathrm{FD},\mathrm{NB}]$, where $x_{\mathrm{TB}\leftarrow\mathrm{FD}} \equiv \frac{k_1}{k_{\mathrm{d}}}[\mathrm{P}]_{\mathrm{FD}}$ and $x_{\mathrm{TB}\leftarrow\mathrm{NB}} \equiv \frac{k_3}{k_{-3}\left([\mathrm{D}_0]_{\mathrm{NB}} - [\mathrm{PD}]_{\mathrm{NB}}\right)}[\mathrm{PD}]_{\mathrm{NB}}$. $[\mathrm{D_o}]_{\mathrm{TB}}$ and $[\mathrm{D_o}]_{\mathrm{NB}}$ are the effective cellular concentrations of TB and NB sites, respectively. Thermodynamic quantities such as the dissociation constant of TB ($K_{\mathrm{d1}} = \frac{k_{\mathrm{o}}^{\mathrm{off}}}{k_1}$) and NB ($K_{\mathrm{d2}} = \frac{k_{-2}}{k_2}$) were also determined from this analysis.

1.5 Supplementary information to chapter 1

1.5.1 Construction of strains and plasmids

All plasmids, primers, and strains used are listed in Supplementary Table 1, Supplementary Table 2, and Supplementary Table 3, respectively. Plasmids and genome extraction were performed using the QIAprep Spin Miniprep Kit (Qiagen), and Wizard Genomic Extraction Kit (Promega), respectively. PCR products as well as digested products including plasmids and inserts were recovered using the Wizard SV Gel and PCR Clean-Up

System (Promega). The primers and all enzymes including restriction enzymes and ligase were purchased from the Integrated DNA Technologies and New England Biolabs, respectively. PCRs were performed using the AccuPrime Pfx DNA Polymerase Kit, and colony PCRs for screening were conducted using the Econo Taq DNA Polymerase Kit (Lucigen).

Tagging E. coli chromosomal genes with mEos3.2-FLAG (i.e., mE) via Lambda-Red

The Zur^{mE} strain was derived from the *Escherichia coli* BW25113 strain (CGSC# 7739 Keio Collection, Yale; genotype: (F-*Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), *λ*-, *rph-1*, *Δ(rhaD-rhaB)568*, *hsdR514*), and was generated by fusing the C-terminus of Zur with the monomeric, irreversibly photoconvertible fluorescent protein mEOS3.2 and FLAG tag for immunoblotting at the *zur* chromosomal locus via Lambda-Red homologous recombination technique.

For the linear DNA insert to tag chromosomal *zur* gene, primer pair H1-EZur-mEOS32-FP and H2-EZur-CAM-RP were used to obtain *H1-mEos3.2FLAG:cat-H2* via PCR from the *pUCmEOS3.2FLAG:cat* template. For the flanking homology regions (H1, and H2), H1 is the last 40 bp of the *zur* gene before the stop codon and H2 is the next 40 bp after the *zur* stop codon.

Transformation of the linear DNA insert into *E.coli* cells was performed via electroporation. We first prepared the electrocompetent *E. coli* BW25113 cells harboring the temperature-sensitive pKD46 plasmid. The SOB media [2% w/v Bacto Tryptone (Sigma-Aldrich, cat. #: T9410), 0.5 % w/v Bacto Yeast Extract (Sigma-Aldrich, cat.#:Y1625), 10 mM NaCl (Macron, 7581-12), 2.5 mM KCl (Fisher Scientific, P217-500), 10 mM $MgCl_2$

(Mallinckrodt, 5958-04), and 10 mM $MgSO_4$ (Fisher Scientific, M63-500) all in nanopure sterile water] containing ampicillin (100 μg/mL, USBiological) and 20 mM L-arabinose (Sigma-Aldrich, cat. #: A3256), which is a reagent that can induce the expression of the *bet*, *gam*, and *exo* λ-Red enzymes encoded in pKD46, was used for culturing. The cells were centrifuged and then washed twice with cold 10% glycerol (Macron, 5092-02) in nanopure sterile water. The cells were diluted to a final volume of 50 μL in 10% glycerol in nanopure sterile water. The linear DNA insert was then electroporated (2.5kV, using MicroPulser Elctroporator; cat.#: 1652100, Bio-Rad) into the prepared electrocompetent cells expressing the recombinase enzymes (exo, β, γ) from pKD46, and then recovered in SOC medium [SOB medium + 20 mM glucose (sigma-Aldrich, cat. #: G7528)]. After 4 hours incubation at 37 ºC, the cells were plated onto LB-agar containing chloramphenicol (10 μg/mL, USBiological), and further incubated for 18 hours.

Successful transformation (i.e., insertion of *mEos3.2FLAG:cat* at the chromosomal target) was confirmed by colony PCR screening and gene sequencing. The temperature-sensitive pKD46 plasmid after homologous recombination was eliminated by incubating the cells at 42 °C for 18 h, which was confirmed by ampicillin-sensitivity. This resulting engineered strain containing *zur-mEos3.2-FLAG-cam* in genome is called Zur^{mE}.

Making the Zur^{mE} gene fusions in the L-arabinose inducible plasmid pBAD24

The gene of *zur-mEos3.2-FLAG* was cloned out from the purified genome of Zur^{mE} as a template, with primer pair EcoRI-EZur-pB24fp and Sal1-FLAG-mEOS32-pB24rp (Supplementary Table 2). The copied PCR product was digested with EcoRI-HF and SalI-HF restriction enzymes and inserted into similarly digested pBAD24 plasmids using Quick Ligase

enzyme to generate pZur_mE (Supplementary Table 1). The plasmids were transformed into the cloning strain E. cloni 10G (Lucigen). Successful transformation was confirmed by antibiotic (100 μg/mL ampicillin) selection on a LB-agar plate, and insertion was screened by colony PCR and confirmed by DNA sequencing. Extracted pZur_mE was transformed into the Δ*zur* strain (JW5714-1 from Keio collection) giving DZ-pZurmE (Supplementary Table 3). The Δ*zur* strain (not Zur^{mE}) was chosen as the base strain for plasmid expression to avoid the complexity that the chromosome and plasmid express different forms of Zur (i.e., Plasmid insert could be a mutant form, e.g., apo mutant or salt-bridge removed form).

Making the apo and salt-bridge removed mutants of Zur in pBAD24

Site-directed mutagenesis was performed using QuikChange mutagenesis kit (Stratagene) to make apo- and salt-bridge removed mutant form of Zur in pBAD24. We mutated a cysteine in the regulatory Zn binding site to serine (C88S) to make the apo-form of Zur, pApoZur_mE, and an aspartate which is responsible for salt-bridge formation between dimers to alanine (D49A) to make the salt-bridge removed mutant of Zur, pZurD49A_mE[15]. Primer pair (EZurC88S-fp and EZurC88S-rp for apo-mutant, and EZurD49A-fp and EZurD49A-rp for salt-bridge removed mutant) was used together with pZur_mE plasmid as the template (Supplementary Table 1, Supplementary Table 2). Sequential mutations were conducted to achieve double mutations (that is, C88S/D49A). Dpn1 digestion (NEB, R0176S) was performed to remove the methylated nonmutated parental plasmid. The plasmid containing the mutant-form of Zur was transformed into E. cloni 10G strain for propagation. Extracted mutant plasmids were confirmed by DNA sequencing. The sequence-confirmed plasmid containing pApoZur_mE, pZurD49A_mE, or pApoZurD49A was transformed into the *zur* strain

JW5714-1, resulting in DZ-pApoZurmE, DZ-pZurD49AmE, and DZ-pApoZurD49AmE strains, respectively (Supplementary Table 3).

Supplementary Table 1. Plasmids used or constructed in this study

Plasmid name	**Relevant characteristics, or insert**	**Source**
pKD46	bet, gam, exo enzymes	Datsenko et al[37]
pBAD24	Base plasmid, L-arabinose inducible	Guzman et al[38]
pUCmEos3.2FLAG:cat	*mEos3.2* : *chloramphenicol*	Chen et al[11]
pZur_mE	*Zur-mEos3.2-FLAG*	This study
pApoZur_mE	*Zur(C88S)-mEos3.2-FLAG*	This study
pZurD49A_mE	*Zur(D49A)-mEos3.2-FLAG*	This study
pApoZurD49A_mE	*Zur(C88S,and D49A)-mEos3.2-FLAG*	This study

Supplementary Table 2. Primers used in this study

Primer name	**5'→3'**
H1-EZur-mEOS32-FP	CCAGCATGATCACTCTGTGCAGGTGAAAAAGAAACCGCGTATGAGTGCGATTAAGCCAGA
H2-EZur-CAM-	TAATCCCTCCTGCCCGACGTGTACAAGGATGTACGCCCTCCG

RP	ACGGCCAGTGAATTCGA
EcoRI-EZur-pB24fp	CAGGAG GAATTCACCATGGAAAAGACCACAACGCA
Sal1-FLAG-mEOS32-pB24rp	AGTCAGGTCGACTTATTTATCATCATCATCTTTATAATCAGGA CGTCGTCTGGCATTGTC
EZurC88S-fp	CAGTTATGTGCTCTCTCATCTGTTCGATC
EZurC88S-rp	GATCGAACAGATGAGAGAGCACATAACTG
EZurD49A-fp	ATGATCTGCTTGCTTTACTGCGCG
EZurD49A-rp	CGCGCAGTAAAGCAAGCAGATCAT

Supplementary Table 3. Strains used or constructed in this study

Strains	**Plasmid**	**Chromosomal Gene modification**	**Reference or source**
BW25113	pKD46	Base strain	Keio collection
Zur^{mE}	none	Zur-mEos3.2-FLAG	This study
JW5714-1	none	Δ*zur*	Keio collection
DZ-pZurmE	pZur_mE	Δ*zur*	This study
DZ-pApoZurmE	pApoZur_mE	Δ*zur*	This study
DZ-pZurD49AmE	pZurD49A_mE	Δ*zur*	This study
DZ-pApoZurD49AmE	pApoZurD49A_mE	Δ*zur*	This study

1.5.2 Intactness and functionality of mE-tagged proteins in E. coli cells

Western blot shows that Zur^{mE} stays intact in the cell

We performed Western blot to check the intactness of the fusion protein (that is, Zur-mEos3.2-FLAG) in the cell. We detected the FLAG epitope (RPDYKDDDDK) at the C-terminal of the fusion protein, and anti-FLAG antibody was used for immunoblotting.

The DZ-pZurmE strain which could express Zur^{mE} from the pBAD24 inducible by L-arabinose, and a negative control strain DZ-pBAD containing the parent pBAD24 without insert were grown overnight (18 hr) in 6 mL LB with appropriate antibiotics. 50 μL of the samples were further grown to OD_{600} of 0.4 in 5 mL M9 medium with amino acids (8% v/v 50x GIBCO), vitamins (4% 100x GIBCO), glycerol (0.4%). L-arabinose was added to the final concentration of 1 mM and the culture was further incubated for 20 min to induce the plasmid expression. For the additional negative controls, samples without L-arabinose induction were prepared. 1 mL aliquots of the resulting cell cultures were collected by centrifugation, and the cell pellets were re-suspended in 200 μL 2X SDS lysis buffer. The lysed samples were run in SDS-PAGE with ECL Plex fluorescent rainbow protein molecular weight markers (GE Healthcare Life Science) in 1X MES buffer, and then transferred onto the Hybond-LEP PVDF membrane (GE Healthcare Life Sciences). The transferred membrane was blocked with 4% Amersham ECL Prime blocking reagent (GE Healthcare Life Sciences) in PBS-T (0.1% Tween-20, Sigma-Aldrich) wash buffer while shaking at RT for 2 hr. After blocking, the membrane was washed with PBS-T twice, and incubated with rabbit-derived anti-FLAG primary antibody (1:10,000 dilution, Rockland Immunochemical) for 2 hr. The membrane was rinsed with PBS-T 4 times and PBS buffer 3 times. The goat-derived Horseradish Peroxidase-

conjugated Fab fragment anti-rabbit antibody (1:5,000 dilution, Rockland Immunochemical) was used as the secondary antibody, which could be probed with Pierce ECL 2 Western Blotting substrate (Fisher Scientific). Bio-Rad ChemiDoc MP Imaging System was used to detect peroxidase activity.

One dominant band of L-arabinose induced Zur-mEos3.2-FLAG was observed at MW ~ 46 kDa (that is, MW of Zur + MW of mEOS3.2-FLAG), and no discernable band was observed at MW ~ 27 kDa which is expected to be the MW of mEOS3.2-FLAG (Supplementary Figure 1a). Therefore, the Zur-mEos3.2-FLAG is intact in the cell.

RT (Reverse Transcription)-PCR shows that Zur^{mE} is functional

RT-PCR assay was performed to measure the mRNA levels of a *zur* regulon, which could directly reflect the functionality (repression ability) of the fusion protein. The *zinT* gene encoding the periplasmic zinc trafficking protein ZinT was chosen as the reference *zur* regulon[14, 15].

The wild-type (BW25113), *zur*^{mE}, Δ*zur*, DZ-pZurmE, and DZ-pApoZurmE strains (Supplementary Table 3) were grown overnight (18 hrs) in 6 mL LB with appropriate antibiotics. 50 μL of the sample were further grown to OD_{600} of 0.4 in 5 mL LB, and then final concentration of 1 mM L-arabinose (and a final 200 μM $ZnSO_4$ if needed for zinc stress) was added to induce (if applicable) the plasmid expression for 30 min. Cells were collected by centrifugation. Total RNA was purified using PureLink® RNA Mini Kit (Life Technologies), and quantified by NanoDrop spectrophotometer (Thermal Scientific) for normalization. Purified RNAs were converted to the cDNA using SuperScript® III First-Strand Synthesis

SuperMix kit (Life Technologies) with random hexamer primers. The primer pair, zinT18-fp and zinT104-rp was used for qPCR[15] (Supplementary Table 2). The mRNA levels of *zinT* were assessed relatively to that of the housekeeping gene (16S rRNA) as an internal reference[49]. The samples were measured in triplicate in an optical 96-well plate (Life Technologies) in a reaction mixture with the SYBR Green reagent (Life Technologies) and imaged using Applied Biosystems (Thermo Fisher) Viia7 Sequence Detection System. Fluorescence was detected at the annealing phase in qPCR. The threshold cycles (C_T) were calculated using Viia7 software.

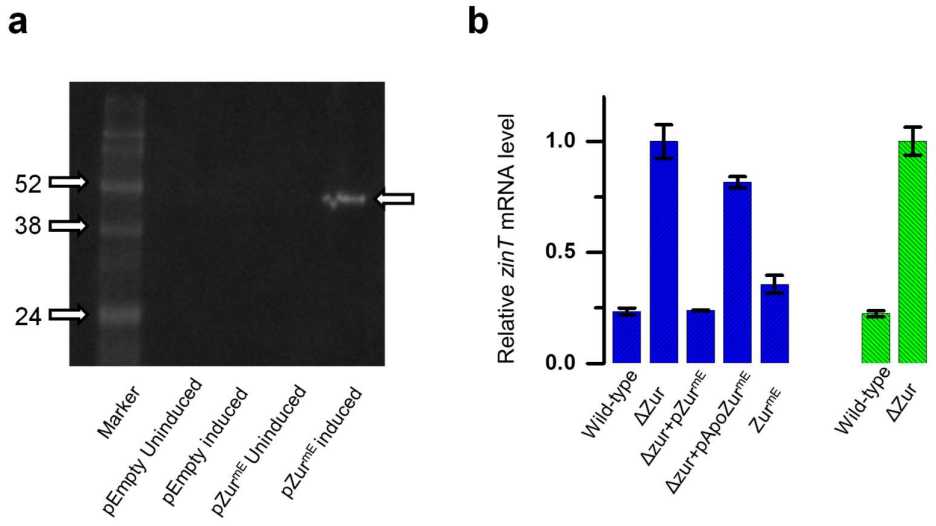
The relative expressions of *zinT* were plotted in Supplementary Figure 1b using the comparative C_T method[50]. The lost repression ability of Δ*zur* was recovered by plasmid expressed Zur^{mE} (Supplementary Figure 1b, 2nd and 3rd column), whereas the apo form of Zur^{mE} (i.e., C88S) could not repress as much as the wild-type strain (Supplementary Figure 1b, 4th column), indicating that the plasmid expressed Zur^{mE} is functional and that the mutation C88S could remove the ability of Zur^{mE} to repress – the latter was expected from the previous report by Gilston et al[15]. The same assay of chromosomally tagged Zur^{mE} strain demonstrated that the tagged-protein naturally expressed from chromosome has ~84% repression compared with the wild-type (Supplementary Figure 1b, 1st and 5th column).

Validation of 20 μM Zn^{2+} as optimal zinc stress condition

In our imaging experiment, we used a concentration of 20 μM Zn^{2+} in the medium as the zinc stress condition, and expected that the Zur proteins in the cell would be predominately in their Zn-metallated form (i.e., holo-form). In this section, we used the RT-PCR assay and zinc quantification method to validate that 20 μM Zn^{2+} could indeed result in sufficient Zn-metallation of the Zur protein in the cell.

We first performed RT-PCR assay under a very high concentration of zinc (that is, 200 μM of Zn^{2+}), and compared the transcript level of *zur* regulon (i.e., *zinT*) to that in the absence of zinc stress in LB. No significant difference was observed between these two conditions, indicating that without addition of extra zinc, the regular LB medium contains a sufficient amount of zinc for Zur to repress its regulon (Supplementary Figure 1b, 1st vs. 6th column).

We then quantified the zinc content of LB medium without addition of extra zinc using a zinc quantitation kit (Zinc Assay kit, Sigma-Aldrich, cat. #: MAK032). The 50 mM of zinc standard solution and its serial dilutions were used for calibration. All samples were treated with TCA solution to free all bound zinc. Samples and standard zinc solution were prepared in the optical 96 plates (Costar) and measured the absorbance at 560 nm using the Synergy HT multi-detection microplate reader (Bio-tek). The zinc content in the regular LB was determined to be ~19.6 μM. Therefore, we concluded that 20 μM Zn^{2+} in the medium is sufficient to metallate the cellular Zur proteins to evoke repression of its regulons.



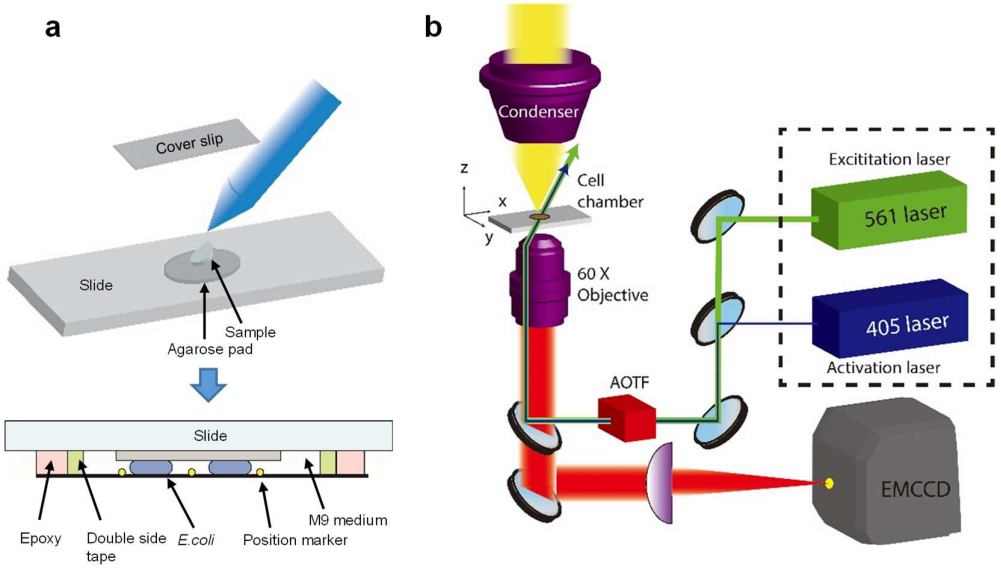
Supplementary Figure 1. Intactness and functionality assay of Zur^{mE}. (**a**) Western blot of

FLAG-tagged Zur^{mE} demonstrated that Zur^{mE} is intact in the cell. Only Zur^{mE} expressed from a pBAD24 plasmid with L-arabinose induction was detectable (5^{th} column). No discernable bands were observed in the negative controls which are the parent empty pBAD24 (without the *zur*^{mE} insert), and un-induced $pZur^{mE}$ (columns 2-4). The expected size of Zur^{mE} is ~ 46 kDa (including the FLAG tag), and no cleavage product was observed (5th column). (**b**) RT-PCR assay for transcription level of *zur* regulon (*zinT*) demonstrated that Zur^{mE} is functional. Columns 1-5 were conducted in 200 μM of Zn^{2+} in LB while columns 6-7 were conducted without zinc stress in LB medium. All mRNA levels were normalized to that of the Δ*zur* strain.

1.5.3 Sample preparation, single-molecule tracking (SMT) via stroboscopic imaging, and Single-cell quantification of protein concentration (SCQPC)

Sample preparation: For imaging experiments, strains were first streaked onto LB plates with appropriate antibiotics. Single colonies were inoculated and grown in LB medium overnight at 37 °C. The overnight cultured was diluted 1:100 in M9 medium supplemented with amino acids, vitamins, and 0.4% glycerol, and further grown to OD_{600} of 0.3. L-arabinose was added to induce plasmid expression for 0 - 20 min when applicable. $ZnSO_4$ solution was also added (for Zn stress) into the cell media to a final concentration of 20 μM. 2 mL of the cell culture was centrifuged and washed twice with the same M9 media, and further incubated at 37 °C for 30 - 60 min to help maturation of the mEos3.2 tag. Cell was collected by centrifugation and then immobilized on an agarose pad between a coverslip and a slide, and

sealed with epoxy-glue (Supplementary Figure 2a). The coverslip was pre-treated with gold particle (100 nm) as position markers for stage-drift correction.



Supplementary Figure 2. Slide preparation for imaging, and schematic diagram of the microscope setup. (**a**) *E. coli* cells were immobilized on 3% agarose gel pad. The sample on the pad was sandwiched between a glass slide and a coverslip which is pre-casted with gold nanoparticles as position markers, and sealed by double-sided tape and epoxy. (**b**) The 405-nm laser was used to photoconvert mEos3.2 to its red-fluorescent form, and the 561-nm laser was used to excite and track the red mEos3.2. An AOTF was synchronized with EMCCD camera. Figure adapted from the reference[11].

SMT and SCQPC. The procedure of SMT via stroboscopic imaging[18-20, 43, 44, 51] and SCQPC were described in Method section and our previous work (Supplementary Figure 2b)[11]. For SMT, the cells were first illuminated with 405-nm laser for 20 ms to photoconvert a single

mE-tagged protein, in which the laser intensity was tuned low enough (1-10 W/cm^2) to photoconvert one or none protein at a time. Then samples were illuminated with 30 pulses of 561-nm laser with 4 ms pulse duration and time lapse T_{tl} = 40 ms to image photoconverted mEos3.2 red fluorescence. The EMCCD camera exposure is slow, but synchronized with the 561 nm pulses. This stroboscopic imaging allowed us to obtain diffract-limited images of mobile molecules, whereas the time lapse was chosen fast enough to follow dynamic protein-DNA interactions. This photoconversion and imaging scheme was repeated for 500 cycles for each cell. An exemplary fluorescence image is shown in Figure 1a.

After the SMT step, SCQPC was carried out. Here the cells were illuminated with 405-nm laser for 2 min to photoconvert all remaining un-photoconverted mEos3.2 proteins to the red form, and the whole cell red fluorescence was imaged using 561-nm laser for 3000 frames at the same power density as the SMT imaging step. This step was repeated twice. The protein copy number inside cells were then determined by dividing the total cell fluorescence by the average fluorescence of a single mEos3.2, which was pre-determined from the earlier SMT. The camera EM-gain was adjusted during the SCQPC step to avoid saturation and remain in the linear dynamic regime of camera sensitivity, and this adjustment was used to correct for the actual recorded camera counts. The total copy number of protein in each cell was estimated using Supplementary Equation (11):

$$N_{copy} = \frac{N_{\mathrm{SMT}} + N_{\mathrm{SCQPC}}}{PE_{\mathrm{mE}}\, OS_{\mathrm{Protein}}} \tag{11}$$

where, PE_{mE} is the photoconversion efficiency of mEos3.2[40, 52] (= 0.42), OS_{protein} is the oligomeric state of protein (Zur is homodimer, therefore $OS_{\mathrm{protein}} = 2$), and N_{SMT} and N_{SCQPC} are

mEos3.2 copy numbers determined in the SMT and SCQPC steps, respectively.

A custom-written MATLAB software called iQPALM (Image-based Quantitative Photo-Activated Localization Microscopy) was used to determine the centroid location of the candidate red fluorescence spots of individual mEos3.2 proteins. We first determined the cell boundary using the bright field optical transmission image. We further selected the cells with length of 2.7 ± 0.9 μm to decrease the possible contamination from dividing cells which potentially have more than one copy of chromosome (Supplementary Figure 3, 2nd column). Then the cell boundaries in the region of interest (ROI) were superimposed onto the corresponding fluorescence image to select out the candidates of single-molecule fluorescence within the cell boundaries. The centroid of each candidate was determined by fitting the fluorescence spot with a 2-dimensional Gaussian function in Supplementary Equation (2).

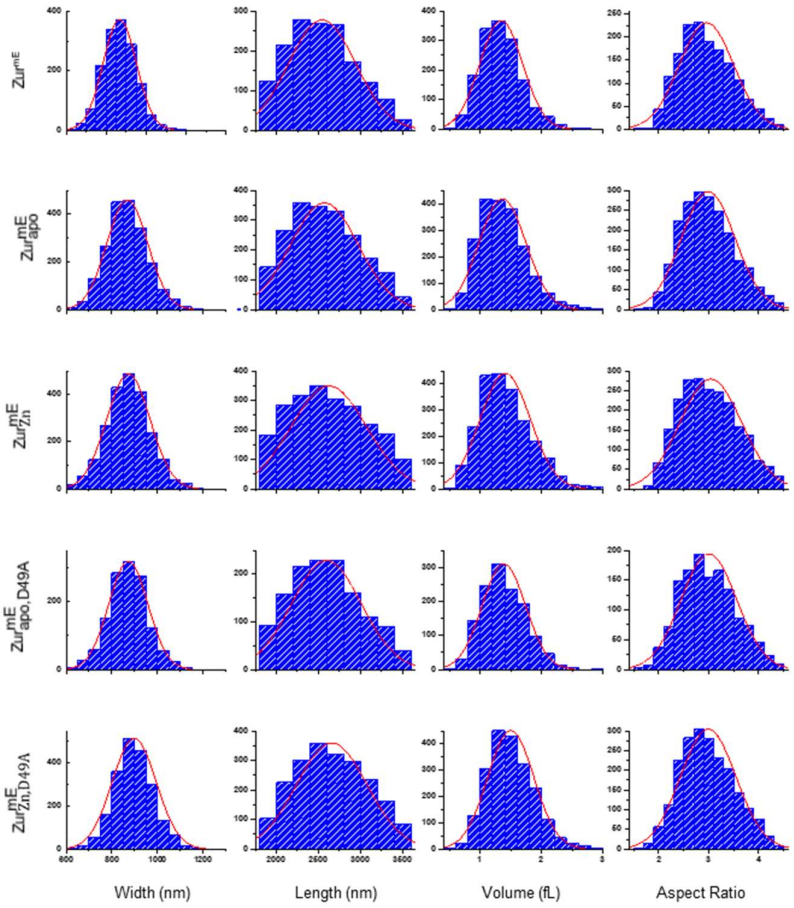
$$I(x,y) = A\exp\left[-\frac{(x-x_0)^2}{2\sigma_x^{\,2}} - \frac{(y-y_0)^2}{2\sigma_y^{\,2}}\right] + B \tag{12}$$

where, $I(x,y)$ is the fluorescence intensity of the candidate at position (x,y), and A, B, (x_o, y_o), and σ_i are the amplitude, background, centroid, and standard deviation of i-direction of the Gaussian function. And the localization precision of its centroid was estimated in Supplementary Equation (13) and (14)[39, 53].

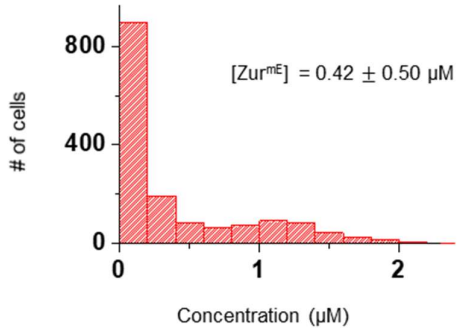
$$Err_i = \sqrt{\frac{\sigma_i^2}{N} + \frac{a^2}{12N} + \frac{8\pi\sigma_i^{\,2}b^2}{a^2N^2}} \tag{13}$$

$$N = \frac{(cts/g)\times(S/QE)\times 3.65}{E_{hv}} \tag{14}$$

Here N is the number of detected photons, a is the pixel size, and b is the standard deviation of the background. *cts, g*, *S*, and *QE* are the total EMCCD counts of the fitted spot, the EM gain (in linear scale, unitless), sensitivity (electrons per count), and quantum yield (unitless) of the EMCCD camera in the spectral range of detected fluorescence, respectively, provided by Andor Technology. The value of 3.65 is a physical constant for electron creation in silicon (eV per electron) and E_{hv} (in eV) is the energy of a single detected fluorescence photon (chosen at wavelength 584 nm, the peak of mEos3.2 red fluorescence spectrum).



Supplementary Figure 3. Distribution of cell geometric parameters. Width and length were obtained from the optical transmission image, and the cell volume was approximated to that of a model geometry in which cell shape in the transmission image was fitted with a cylinder with two hemispherical caps as described in our previous work[11]. Each histogram is fitted with a Gaussian function (red curve).



Supplementary Figure 4. An example of the experimentally sampled distribution of the Zur^{mE} protein concentration among individual cells expressed from a pBAD24 plasmid. Expression level could be controlled by L-arabinose induction level. More experiments were performed on cells having lower protein concentrations so as to achieve sufficient statistics (the number of trajectories per cell obtained is less for cells having lower cellular protein concentrations). The average concentration with standard deviation is denoted in the panel.

1.5.4 Determination of the minimal number of diffusion states and their fractional populations

Probability density function (PDF) and cumulative distribution function (CDF) of displacement length r per time-lapse

The probability density function of displacement vector $\vec{x}$ at time *t* for *N*-dimensional Brownian diffusion is:

$$P_N(\vec{x},t)\mathrm{d}\vec{x} = \left(\sqrt{\frac{1}{4\pi Dt}}\right)^N \exp\left(-\frac{\vec{x}_i^{\,2}}{4Dt}\right)\mathrm{d}\vec{x} \tag{15}$$

where *D* is the diffusion constant. Then, the probability density function of the scalar displacement length *r*, PDF(*r*, *t*), is obtained by integrating $P_N(\vec{x},t)$ over all angular spaces (that is; $\mathrm{d}x\mathrm{d}y \Rightarrow 2\pi r\mathrm{d}r$ for 2D, and $\mathrm{d}x\mathrm{d}y\mathrm{d}z \Rightarrow 4\pi r^2\mathrm{d}r$ for 3D).

$$\mathrm{PDF}_{2\mathrm{D}}(r,t) = \frac{r}{2Dt}\exp\left(-\frac{r^2}{4Dt}\right) \tag{16}$$

$$\mathrm{PDF}_{3\mathrm{D}}(r,t) = \frac{r^2}{Dt\sqrt{4\pi Dt}}\exp\left(-\frac{r^2}{4Dt}\right) \tag{17}$$

The corresponding cumulative distribution function (CDF) of displacement length *r* is obtained by $\mathrm{CDF}(r,t) = \int_0^r \mathrm{PDF}(l,t)\mathrm{d}l$.

$$\mathrm{CDF}_{2\mathrm{D}}(r,t) = 1 - \exp\left(-\frac{r^2}{4Dt}\right) \tag{18}$$

$$\mathrm{CDF}_{3\mathrm{D}}(r,t) = \mathrm{erf}\left(\sqrt{\frac{r^2}{4Dt}}\right) - \sqrt{\frac{r^2}{\pi Dt}}\exp\left(-\frac{r^2}{4Dt}\right) \quad (19)$$

Fitting the histogram of displacement length r with PDF(r, t) or the cumulative distribution of r CDF(r, t) will both give the diffusion constant D. However, there is a practical advantage in using CDF than PDF. Using PDF analysis, the choice of the bin size in generating the histogram of r needs to be carefully evaluated, as it may affect the fitting outcome, while the cumulative distribution does not involve binning. Therefore, we chose to fit the CDF rather than PDF. Also, since we tracked the molecules in 2-D plane, we used the 2-D versions of PDF and CDF for next subsequent data analysis (e.g. Supplementary Equations (16) and (18)).

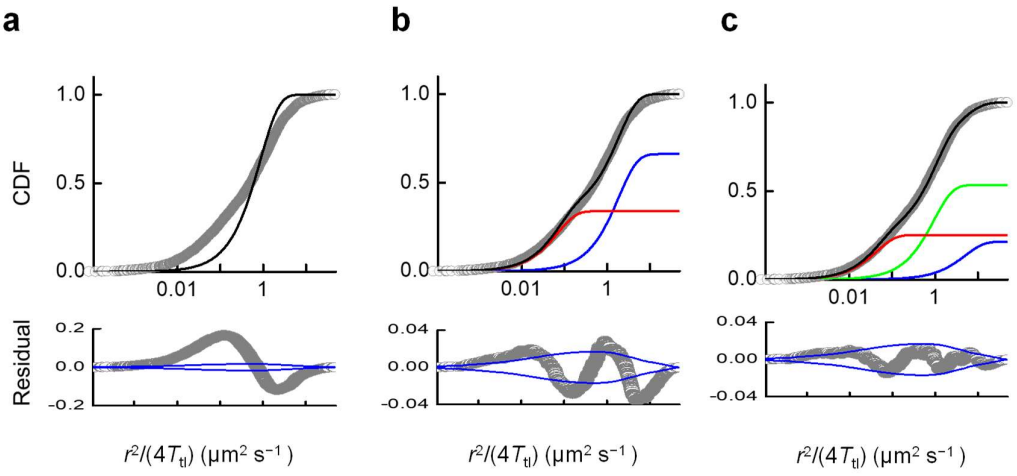
Therefore, the effective diffusion constants and the fractional populations of corresponding states were extracted by analyzing CDF of displacement length r per time-lapse (T_{tl} = 40 ms). Experimentally, only the first displacement of each tracking trajectory was used for constructing CDF to avoid potentially biasing the sampling toward relatively long trajectories[42].

Fits were performed using least-square fitting method in MATLAB, and quality of the fit was assessed by residual analysis. A single component model could not satisfactorily fit the data (Supplementary Figure 5a), indicating more than one diffusion state are present. We therefore used a linear combination of multiple diffusive terms for the CDF(r) (Supplementary equation (20)) and fitted the data until the residual falls within a satisfactory range (i.e., within 95% confidence level). Note that linear combination of CDFs assumes a quasi-static system as an approximation, which means the time-resolution of experimental observation should be fast enough to sample each diffusion state (i.e., sampling is faster than the interconversion rates

between the states). We later obtain interconversion rates between states and verify this approximation (Supplementary Section 0).

$$\mathrm{CDF}(r) = \sum_i A_i \left(1 - \exp\left(-\frac{r^2}{4D_i T_{tl}}\right)\right) \tag{20}$$

Here, D_i and A_i are the diffusion constant of i-th state and its fractional population, respectively, and $\sum_i A_i = 1$. The residual analysis of CDF indicated that three components in Supplementary Equation (20) were needed to satisfactorily fit the data, which means that three diffusion states are *minimally* necessary to characterize the mobility of Zur inside a cell (Supplementary Figure 5c).



Supplementary Figure 5. Upper panels are the CDFs of r (plotted against $\frac{r^2}{4T_{tl}}$) of $\mathrm{Zur}^{\mathrm{mE}}_{\mathrm{apo}}$ at $\left[\mathrm{Zur}^{\mathrm{mE}}_{\mathrm{apo}}\right]_{\mathbf{cell}}$=174 ± 14 nM fitted with (**a**) one, (**b**) two, and (**c**) three diffusion-state models. The grey circles are the experimental CDF data. The colored lines are individual components of CDF(r). The black line is the overall fit. Lower panels are residues of the CDF fits. Blue dashed lines are the 95% confidence bounds and the grey circles are the residuals.

Assignments and validation of the three diffusion states of Zur.

After sorting the individual cells into groups of similar range of cellular protein concentrations, we obtained the CDF(r) for each group and performed global CDF analysis across the cellular protein concentrations, in which the diffusion constants of respective diffusion states were shared while the factional populations were varied, since the effective diffusion constants are expected to be concentration independent. This global analysis of CDF was applied to all strains and conditions. Three terms in CDF(r) were always the minimal number of diffusion states to satisfactorily fit the CDF. Exemplary CDF fits over all cellular protein concentration ranges are shown in Supplementary Figure 6, and the fitted D's and A's are summarized in Supplementary Table 4 and Supplementary Table 5.

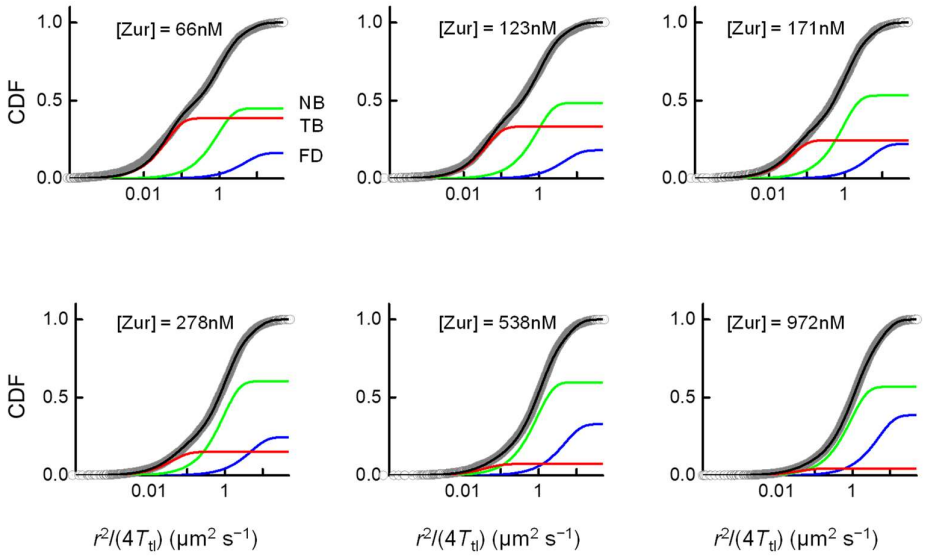
The fastest diffusing state was assign as the FD state (i.e., freely diffusing Zur proteins in the cytosol, D_{FD}), the medium diffusing state as the NB state (i.e., nonspecifically bound to DNA, D_{NB}), and the slowest one as TB (i.e., tightly bound at a chromosomal site, either a consensus recognition site or a tight-binding site that does not have a consensus sequence, D_{TB}) whose motions reflect mostly the chromosome dynamics and localization uncertainties[18-20, 22, 43, 54]. Then Supplementary Equation (20) became:

$$\mathrm{CDF}(r) = A_{\mathrm{FD}}\left(1-\exp\left(-\frac{r^2}{4D_{\mathrm{FD}}T_{\mathrm{tl}}}\right)\right) + A_{\mathrm{NB}}\left(1-\exp\left(-\frac{r^2}{4D_{\mathrm{NB}}T_{\mathrm{tl}}}\right)\right) + (1-A_{\mathrm{FD}}-A_{\mathrm{NB}})\left(1-\exp\left(-\frac{r^2}{4D_{\mathrm{TB}}T_{\mathrm{tl}}}\right)\right) \quad (21)$$

Note that, we assigned the slowest diffusion state as 'tight binding' instead of the term 'specific binding' used in a previous study[11], since the interaction of apo-protein to the Zur box

consensus sequence DNA is known to be much weaker than holo-protein and therefore we postulated that the tight-binding of apo-Zur to chromosome in the cell should be at a non-consensus DNA site. Moreover, we couldn't rule out the possibility that holo-protein could also bind at these non-consensus sequence sites. Therefore, for holo-Zur, tight binding sites may include both consensus and non-consensus sites, and for apo-Zur, tight binding sites only refer to the non-consensus sites.

The extracted effective diffusion constants, D_{FD} = 5.0 ± 0.5, D_{NB} = 0.82 ± 0.05, and D_{TB} = 0.040 ± 0.003 from the CDF analysis were closed to the reported values (D_{FD} = 3.7 ± 0.2, D_{NB} = 0.70 ± 0.03, and D_{SB} = 0.036 ± 0.009) of CueR and ZntR, two metalloregulators in *E. coli*[11] (all units are $\mu m^2/s$). The major difference in FD state stemmed from the time-lapse difference, which more significantly influences the effective diffusion constant of faster moving ones (here T_{tl} = 40 ms, but T_{tl} = 60 ms for the work on CueR and ZntR[11]). Note that the effective diffusion constant of free mEos3.2 is 11.4 ± 0.3 at T_{tl} = 15 ms. Also, due to the crowded environment in the cytosol, the time-lapse dependence of D follows approximately a power law fashion (that is, $D = D_\alpha T_{tl}^\alpha$)[11, 55]. These previous observations could provide the expected range of D_{FD}, ($4.5 \leq D_{FD} \leq 6.5$), which directly validate our FD state assignment.



Supplementary Figure 6. Exemplary global analysis of protein-concentration-sorted CDF of displacement length r (plotted against $r^2/(4T_{tl})$) for Zur^{mE}. The grey circles are the experimental CDF data; the black curve is the overall fit of three diffusion states of CDF(r), which are individually plotted as blue, green, and red curves for FD, NB, and TB states, respectively. The mean values of total cellular protein concentration [Zur^{mE}] are given in the respective panels.

Supplementary Table 4. Extracted effective diffusion constants of Zur variants in live *E. coli.* cells

	D_{FD} (μm^2/s)	D_{NB} (μm^2/s)	D_{TB} (μm^2/s)
Zur^{mE}	4.93 ± 0.59	0.92 ± 0.07	0.0395 ± 0.0044
Zur$^{mE}_{apo}$	5.01 ± 0.46	0.82 ± 0.05	0.0396 ± 0.0034
Zur$^{mE}_{Zn}$	6.65 ± 0.51	0.82 ± 0.05	0.0327 ± 0.0028

$Zur^{mE}_{apo,\ D49A}$	5.47 ± 0.96	0.82 ± 0.06	0.047 ± 0.008
$Zur^{mE}_{Zn,\ D49A}$	5.83 ± 0.60	0.85 ± 0.05	0.044 ± 0.008

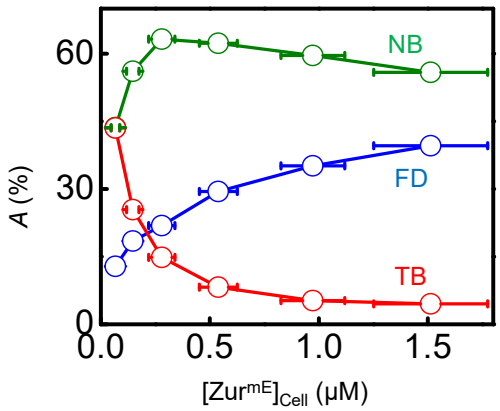
Supplementary Table 5. Extracted fractional populations of Zur variants in live *E. coli.* cells

	$[P]_{cell}$ (nM)	A_{FD} (%)	A_{NB} (%)	A_{TB} (%)
Zur^{mE}	66 ± 19	17.6 ± 0.3	44.2 ± 0.4	38.2 ± 0.4
	123 ± 15	18.3 ± 0.4	48.8 ± 0.4	32.9 ± 0.5
	171 ± 14	25.6 ± 0.6	53.5 ± 0.9	20.9 ± 0.7
	279 ± 61	24.0 ± 0.5	62.5 ± 0.6	13.5 ± 0.3
	538 ± 87	33.3 ± 0.6	58.6 ± 0.5	8.1 ± 0.3
	973 ± 146	39.8 ± 0.4	55.9 ± 0.4	4.3 ± 0.2
	1508 ± 253	42.6 ± 0.5	53.6 ± 0.4	3.8 ± 0.1
Zur^{mE}_{apo}	63 ± 19	18.0 ± 0.3	50.3 ± 0.5	31.7 ± 0.4
	124 ± 15	21.7 ± 0.4	48.2 ± 0.6	30.1 ± 0.5
	174 ± 14	19.6 ± 0.6	56.9 ± 0.4	23.5 ± 0.7
	287 ± 57	27.7 ± 0.6	57.6 ± 0.3	14.7 ± 0.4
	527 ± 90	38.0 ± 0.4	52.5 ± 0.5	9.5 ± 0.4
	936 ± 142	41.5 ± 0.3	53.9 ± 0.4	4.6 ± 0.3
	1589 ± 287	43.4 ± 0.2	50.6 ± 0.3	5.9 ± 0.1
Zur^{mE}_{Zn}	62 ± 21	12.8 ± 0.2	49.6 ± 0.2	37.6 ± 0.4

	122 ± 14	18.1 ± 0.3	47.2 ± 0.6	34.7 ± 0.6
	174 ± 15	22.8 ± 0.3	42.9 ± 0.4	34.3 ± 0.5
	296 ± 59	36.1 ± 0.4	46.1 ± 0.4	17.8 ± 0.4
	528 ± 86	39.8 ± 0.4	50.9 ± 0.3	9.3 ± 0.2
	944 ± 141	44.3 ± 0.2	48.6 ± 0.2	7.1 ± 0.1
	1521 ± 223	42.7 ± 0.5	51.1 ± 0.4	6.2 ± 0.2
	69 ± 17	15.1 ± 0.3	54.6 ± 0.6	30.3 ± 0.6
	119 ± 14	19.5 ± 0.5	55.8 ± 0.6	24.7 ± 0.6
	175 ± 15	17.0 ± 0.8	63.5 ± 0.8	19.5 ± 0.3
$Zur^{mE}_{apo,\ D49A}$	295 ± 56	25.2 ± 0.6	66.3 ± 0.6	8.5 ± 0.2
	525 ± 86	29.4 ± 0.7	63.8 ± 0.5	6.8 ± 0.3
	912 ± 151	35.9 ± 0.6	58.7 ± 0.5	5.4 ± 0.3
	1321 ± 87	31.6 ± 0.9	63.4 ± 1.2	5.0 ± 0.5
	71 ± 16	13.3 ± 0.4	54.2 ± 0.5	32.5 ± 0.3
	120 ± 15	20.5 ± 0.4	52.3 ± 0.5	27.2 ± 0.5
	172 ± 15	20.9 ± 0.5	58.0 ± 0.6	21.1 ± 0.8
$Zur^{mE}_{Zn,\ D49A}$	302 ± 56	24.6 ± 0.4	57.0 ± 0.5	18.4 ± 0.5
	537 ± 88	27.3 ± 0.4	61.8 ± 0.2	10.9 ± 0.3
	930 ± 148	31.8 ± 0.3	61.2 ± 0.2	7.0 ± 0.2
	1448 ± 192	32.5 ± 0.2	63.4 ± 0.2	4.1 ± 0.1

Protein concentration dependence of fractional populations (A_{FD}, A_{NB}, and A_{TB}).

The resolved CDF also provided fractional populations (that is A_{FD}, A_{NB}, and A_{TB}) of corresponding diffusion states. Exemplary protein-concentration-dependent fractional populations are shown in Figure 1d and Supplementary Figure 7. Complete results are listed in Supplementary Table 5.



Supplementary Figure 7. Exemplary protein-concentration-dependence of the fractional populations of Zur^{mE} strain diffusion states. The FD, NB, and TB states are plotted as blue, green, and red symbols, respectively. Error bars of protein concentrations are s.d.

With increasing cellular protein concentration, the fractional population of FD state, A_{FD}, increases while that of TB, A_{TB} decreases. These trends could further validate their assignments: at high protein concentrations, each protein molecule will spend more time freely diffusing than tightly bound to TB sites on chromosome since more protein molecules compete for the limited number of TB sites. On the other hand, the fraction of TB state out of the total

protein should decrease with increasing concentration, again because there is limited number of TB sites on DNA. At the very high cellular protein concentrations, A_{TB} becomes negligible (< 5%), where we later extracted k_{NB}, the unbinding rate constant from NB sites (see Supplementary Section 1.5.5).

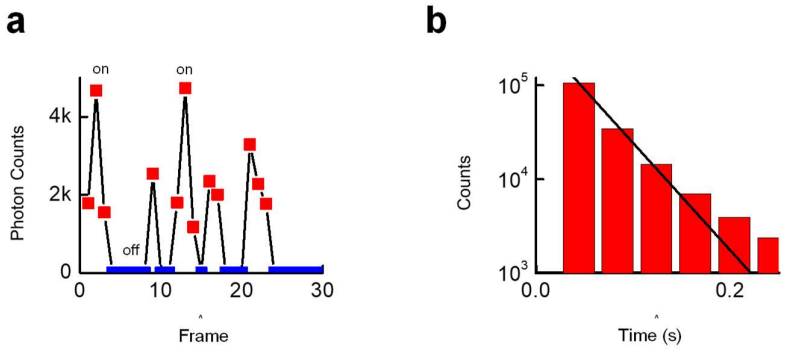
1.5.5 Determination of apparent unbinding rate constant k_d from TB site

The apparent unbinding rate constant k_d of Zur was determined by analyzing the distributions of the microscopic residence time τ at TB sites, extracted from single-molecule displacement-vs-time trajectories, as we[11, 56] have reported. When Zur is bound to a site on the chromosome tightly, its motion is stalled, which is reflected by the corresponding small displacement length r. Therefore, we could obtain the microscopic residence time τ by thresholding the displacement-vs-time trajectory with an upper limit r_o (=200 nm), which was chosen to include >99.5% of the TB state, as reflected in the PDF(r, t) (Figure 1b); a complete τ should start upon transition from a large r to below r_o, and terminate upon transitions to above r_o.

Experimentally, the extracted residence time τ by r_o-thresholding could also be terminated by photobleaching/blinking of the mE tag. The photobleaching/blinking rate constant k_{bl} was independently determined from analyzing the distribution of the fluorescence on-times in the tracking trajectories (Supplementary Figure 8a). Since we used stroboscopic imaging in which the 561 nm excitation laser was only illuminating during the integration time T_{int} (4 ms), and remained dark for the rest time within each time-lapse $\underline{T}_{tl}$ (40 ms), the apparent decay constant from the fluorescence on-time distribution needed to be corrected by the ratio of T_{int} and $\underline{T}_{tl}$. Therefore, the distribution of on-time was fitted with the equation below.

$$f_{on}(t) = N\exp\left(-k_{\mathrm{bl}}\frac{T_{\mathrm{int}}}{T_{\mathrm{tl}}}t\right) \quad (22)$$

N is a normalization constant. An exemplary fluorescence intensity trajectory and the histogram of the fluorescence on-times are shown in Supplementary Figure 8a and b. The extracted k_{bl} is 263.2 ± 11.5 s^{-1}, consistent with the reported value (257 ± 9 s^{-1}) under similar 561 nm excitation conditions[11].



Supplementary Figure 8. Determination of photobleaching/blinking rate constant k_{bl}. (**a**) An example of single-molecule mE fluorescence vs. time trajectory of $\mathrm{Zur}^{\mathrm{mE}}_{\mathrm{apo}}$ in one imaging cycle. Red/blue square represents on/off time. The on time frames are first identified by their pixel counts to be above the threshold which is the mean value plus four standard deviation of the pixel counts of the whole image; then the identified on frames were analyzed by the PSF fitting procedure to get the integrated fluorescence intensity of each fitted PSF, as described in Supplementary Section 1.5.3. One frame = 40 ms. (**b**) The distribution of fluorescence on-time for $\mathrm{Zur}^{\mathrm{mE}}_{\mathrm{apo}}$ (red bar) and the fitting with Supplementary Equation (22) (black curve) to obtain k_{bl}.

Diffusion process is probabilistic at the microscopic level, however. Even proteins freely diffusing (FD) or non-specifically bound (NB) to chromosome, which have relatively large effective diffusion constant (5.0 μm^2/s and 0.82 μm^2/s, respectively, for Zur^{mE}_{apo}), have finite probabilities to have small displacement length (e.g., Figure 1b), and thus will contribute to the distribution of residence time τ thresholded by r_o. The proportions of FD and NB states below r_o (= 200 nm) are 4.9%, and 26.3%, respectively (e.g., PDF(r) in Figure 1b), as compared with that >99.5% of the TB state. To deconvolute the contributions from FD and NB states to the residence time distribution, we first calculated the survival probability $S(r_o, t)$, which is the probability for a protein that was at the origin to survive within a circle of radius r_o within time t, as we previously derived[11]. This probability is a product of multiple terms, each representing an independent process that affects this survival, including diffusion, dissociation of protein from chromosome, and photobleaching/blinking of mE (Supplementary Equation (23)).

$$S(r_o,t)=\left[1-\exp\left(-\frac{r_o^2}{4Dt}\right)\right]\exp(-k_{eff}t) \tag{23}$$

Here, the term in [ ] corresponds to the survival probability due to diffusion, and k_{eff} is the sum of the unbinding rate constant (applicable for the TB and the NB state only) and the effective photobleaching/blinking rate constant (i.e., $k_{bl}\frac{T_{int}}{T_{tl}}$). The overall survival probability within r_0 for a Zur protein is a linear combination of survival probabilities of each state weighted by its fractional population.

$$S_{all}(r_o,t)=A_{FD}S_{FD}(r_o,t)+A_{NB}S_{NB}(r_o,t)+A_{TB}S_{TB}(r_o,t) \tag{24}$$

Then, the respective probability distribution function of the thresholded residence time τ, for the FD, NB, and TB states (that is, $\varphi_{FD}(\tau)$, $\varphi_{NB}(\tau)$, and $\varphi_{TB}(\tau)$) can be obtained by taking

a time-derivative of the survival probability (i.e., $\varphi(\tau)=-\frac{\partial S(t)}{\partial t}|_{t=\tau}$):

$$\varphi_{\mathrm{all}}(\tau) = A_{\mathrm{FD}}\varphi_{\mathrm{FD}}(\tau)+A_{\mathrm{NB}}\varphi_{\mathrm{NB}}(\tau)+A_{\mathrm{TB}}\varphi_{\mathrm{TB}}(\tau) \tag{25}$$

$$\varphi_{\mathrm{FD}}(\tau)=\left[\frac{r_{\mathrm{o}}^2}{4D_{\mathrm{FD}}\tau^2}\exp\left(-\frac{r_{\mathrm{o}}^2}{4D_{\mathrm{FD}}\tau}\right)+k_{\mathrm{eff}}^{\mathrm{FD}}\left(1-\exp\left(-\frac{r_{\mathrm{o}}^2}{4D_{\mathrm{FD}}\tau}\right)\right)\right]\exp\left(-k_{\mathrm{eff}}^{\mathrm{FD}}\tau\right) \tag{26}$$

$$\varphi_{\mathrm{NB}}(\tau)=\left[\frac{r_{\mathrm{o}}^2}{4D_{\mathrm{NB}}\tau^2}\exp\left(-\frac{r_{\mathrm{o}}^2}{4D_{\mathrm{NB}}\tau}\right)+k_{\mathrm{eff}}^{\mathrm{NB}}\left(1-\exp\left(-\frac{r_{\mathrm{o}}^2}{4D_{\mathrm{NB}}\tau}\right)\right)\right]\exp\left(-k_{\mathrm{eff}}^{\mathrm{NB}}\tau\right) \tag{27}$$

$$\varphi_{\mathrm{TB}}(\tau)=k_{\mathrm{eff}}^{\mathrm{TB}}\exp\left(-k_{\mathrm{eff}}^{\mathrm{TB}}\tau\right) \tag{28}$$

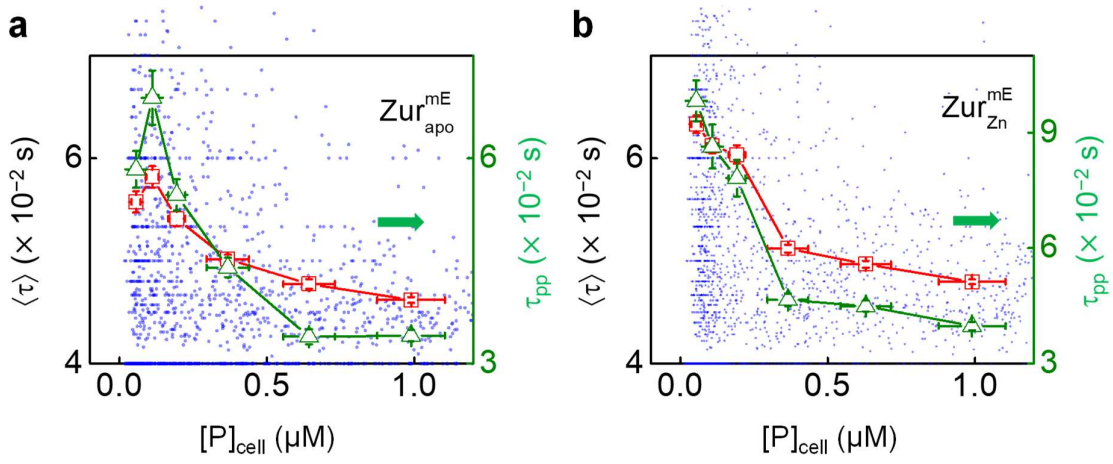
Here, $k_{\mathrm{eff}}^{\mathrm{FD}} = k_{\mathrm{bl}}\frac{T_{\mathrm{int}}}{T_{\mathrm{tl}}}$, $k_{\mathrm{eff}}^{\mathrm{ND}} = k_{\mathrm{bl}}\frac{T_{\mathrm{int}}}{T_{\mathrm{tl}}} + k_{\text{-}2}$ and $k_{\mathrm{eff}}^{\mathrm{TB}} = k_{\mathrm{bl}}\frac{T_{\mathrm{int}}}{T_{\mathrm{tl}}} + k_{\mathrm{d}}$ (see Figure 2c for definition of rate constants). Note, since r_{o} (=200 nm) was chosen to include >99.5% displacement length of the TB state, the time-derivative of the diffusion-term in $S_{\mathrm{TB}}(r_{\mathrm{o}},t)$ became negligible. Therefore, only the exponential term survived for the TB state (that is, $S_{\mathrm{TB}}(r_{\mathrm{o}},t) = \exp\left(-k_{\mathrm{eff}}^{\mathrm{TB}}t\right)$). $k_{\text{-}2}$ is the unbinding rate constant from the NB sites. $k_{\text{-}2}$ was extracted from the highest cellular concentration regime by fitting the residence time distribution with Supplementary Equation (29), which is a reduced version of Supplementary 16 in which A_{TB} is <5% and $A_{\mathrm{TB}}\varphi_{\mathrm{TB}}(\tau)$ term becomes negligible:

$$\varphi_{all}(\tau) = A_{\mathrm{FD}}\varphi_{\mathrm{FD}}(\tau)+A_{\mathrm{NB}}\varphi_{\mathrm{NB}}(\tau) \tag{29}$$

k_{d} was extracted by fitting the residence time distribution with Supplementary Equation (25) with predetermined D's, A's, k_{bl}, and $k_{\text{-}2}$. All determined rate constants are

summarized in Supplementary Table 6. This method of extracting k_d from residence time distribution was rigorously verified by using simulation data, in our previous study[11].

We further obtained the simple average of the residence time $\langle \tau \rangle$, and the single exponential fit of the residence time distribution (i.e., fitting the distribution with $y = \exp\left[-\left(\frac{1}{\tau_{PP}} + k_{bl} \frac{T_{int}}{T_{tl}} \right) t \right]$, which includes the time constant τ_{PP}, for which $1/\tau_{pp}$ approximates the unbinding rate constant, and the correction for photobleaching/blinking process) for each group of cells with similar cellular protein concentrations. Since $\langle \tau \rangle$ and τ_{PP} are inversely related to the apparent unbinding rate constant k_d, the dependence of $\langle \tau \rangle$ and τ_{PP} on cellular protein concentration might have a biphasic behavior (increasing-followed-by-decreasing) for Zur^{mE}_{apo} with increasing cellular protein concentrations, or a monotonous decreasing behavior for Zur^{mE}_{Zn}. Supplementary Figure 9 clearly showed the expected behaviors, supporting the robustness of the biphasic unbinding behavior of Zur from DNA (or the facilitated unbinding of holo-Zur within the accessible protein concentration range).

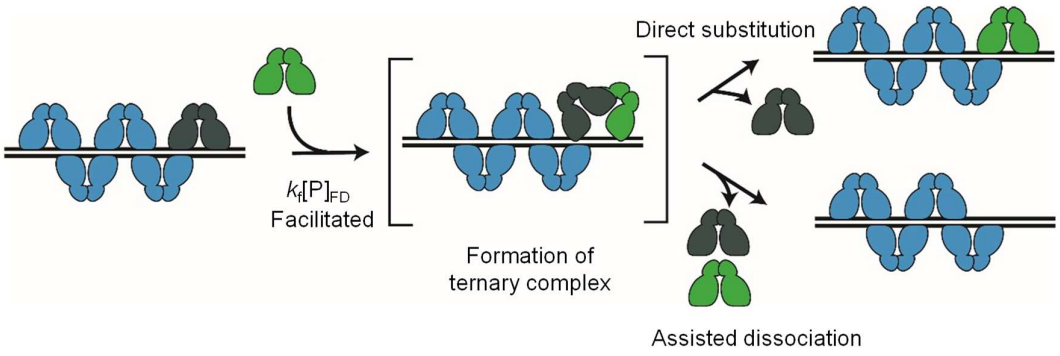


Supplementary Figure 9. Correlation of $\langle \tau \rangle$ with average protein concentration in each cell for Zur^{mE}_{apo} (**a**), and Zur^{mE}_{Zn} (**b**) (each blue dot = 1 cell) from a total of ~1300 cells each. The individual cells are then grouped by similar protein concentration ranges and averaged within each group to obtain the averages (open red squares connected by red lines). Further correction of photobleaching/blinking kinetics of the mE tag gives the corresponding constant τ_{PP} from a single exponential fit (open green triangles connected by green lines). *x, y* error bars are s.d. and s.e.m.

1.5.6 Mechanistic model of the biphasic concentration dependence of k_d: repressed followed by facilitated unbinding

Using single-molecule tracking, our previous study showed that the apparent first-order unbinding rate constants of the metalloregulator CueR and ZntR show linear dependences on their cellular free protein concentration (i.e., facilitated unbinding), highly analogous to the facilitated unbinding phase of Zur observed here. For CueR/ZntR, we proposed that the

facilitated unbinding results from the formation of a ternary complex intermediate, in which two protein dimers are bound to the recognition site, each of which uses one DNA-binding domain attaching to half of the dyad recognition sequence, leading to assisted dissociation and direct substitution eventually[10, 11] (Supplementary Figure 10). Similar facilitated unbinding (i.e., concentration-enhanced) was also observed for the nonspecific dsDNA binding protein nucleoid associate proteins, ssDNA binding protein replication protein A, and DNA polymerases[6-9], and the same mechanism involving a ternary complex was invoked, reaching a mechanistic consensus as we recently reviewed[5]. This ternary complex was also kinetically resolved for CueR interacting with an engineered DNA structure in a previous study of ours[57]. We think the facilitated unbinding phase of Zur observed here also follows the same mechanism (Supplementary Figure 10), for which the effective first-order unbinding rate constant can be written as $k_f[P]_{FD}$, where k_f is a 2nd-order kinetic rate constant and $[P]_{FD}$ is the freely diffusing Zur concentration in the cytosol.

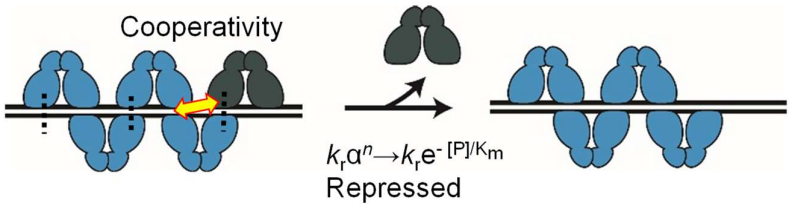


Supplementary Figure 10. Proposed mechanism for facilitated unbinding involving a ternary complex as an intermediate, in which its formation rate is linearly proportional to free-protein concentration in the cytosol. Then it proceeds to direct substitution or assisted dissociation pathway, similarly as we[10, 11] and others[6, 7] described previously. The scheme shown here is

replicated in Figure 2e, lower, in the main text.

Cooperativity among oligomerized proteins can explain the repressed unbinding

In this section, we formulate a quantitative mechanistic model to account for the repressed unbinding behavior of Zur. The crystal structure of *E. coli* Zur (in holo form) on a short 31 bp DNA oligo was solved[15], which showed that two Zur dimers can bind simultaneously to DNA, in which a pair of salt bridges were identified between the two dimers. Later, footprint assay showed Zur from *Streptomyces coelicolor* can oligomerize on DNA, in which the oligomerization number is likely significantly greater than two[25] (note that oligomerization number of the crystal structure of *E. coli* Zur on DNA is 2, i.e., two dimers). Therefore, it is reasonable to postulate that Zur in *E. coli* also has strong oligomerization tendency, i.e., the oligomerization number n_o can be greater than 2, considering that there is ample space on the chromosome. Here we propose that unbinding of Zur is attenuated due to the intermolecular interactions between the dimers that oligomerize at a TB site, as shown schematically in Supplementary Figure 11.



Supplementary Figure 11. Proposed mechanism for repressed unbinding involving oligomerization at a TB site, where the oligomerized proteins have a positive cooperativity via

salt-bridge interactions (black dash lines) for stabilization to make unbinding of a protein slow down. The unbinding rate constant is attenuated exponentially by the number of proteins in the oligomer (Supplementary Equation (31)), leads to an exponential decay dependence on the free-protein concentration in the cytosol (Supplementary Equation (31)). The scheme shown here is replicated in Figure 2e, upper, in the main text.

Based on the proposed mechanism in Supplementary Figure 11, the microscopic kinetic scheme of protein binding-unbinding at any single TB site can be written as:

$$P_{n-1}D + P \underset{k_{\mathrm{d}}^{(n)}}{\overset{k_1}{\rightleftharpoons}} P_nD \tag{30}$$

where, the binding of a dimeric protein P to a TB site that already has n-1 proteins bound on DNA occur with a rate constant k_1, and the corresponding unbinding rate constant is $k_{\mathrm{d}}^{(n)}$, which depends on n, the number of proteins bound at the TB site, which can vary from 1 to the oligomerization number n_0.

To account for the repressed unbinding kinetics, we model that a component of microscopic unbinding rate constant $k_{\mathrm{d}}^{(n)}$ is attenuated by a factor of α^n, where $0 < \alpha < 1$. Along with the facilitated unbinding component $k_{\mathrm{f}}[\mathrm{P}]_{\mathrm{FD}}$ and an intrinsic unbinding component k_0, the overall microscopic unbinding rate constant at a TB site can be written as.

$$k_{\mathrm{d}}^{(n)} = k_{\mathrm{o}} + k_{\mathrm{r}}\alpha^n + k_{\mathrm{f}}[\mathrm{P}]_{\mathrm{FD}} \tag{31}$$

If we assume that the (n-1)'th binding event does not affect n'th binding event

(which is consistent with our model that the binding rate constant k_1 is independent of n) and the dependence of the unbinding of n'th protein on the prior n-1 proteins is a small perturbation, the distribution of n will follow Poisson distribution $\Phi(n)=\frac{\langle n\rangle^n}{n!}\exp[-\langle n\rangle]$, w here $\langle n\rangle$ is the mean value of n.

Then the observable unbinding rate constant k_d is given by

$$\begin{aligned} k_d &= \sum_{n=0}^{n_0}\Phi(n)k_d^{(n)} = k_o + k_r e^{-\langle n\rangle}\sum_{n=0}^{n_0}\frac{\left(\alpha\langle n\rangle\right)^n}{n!} + k_f[P]_{FD} \\ &\simeq k_o + k_r e^{-(1-\alpha)\langle n\rangle} + k_f[P]_{FD} \end{aligned} \tag{32}$$

Note that if n_0 is very large or $\alpha\langle n\rangle$ is very small, $\sum_{n=0}^{n_0}\frac{\left(\alpha\langle n\rangle\right)^n}{n!}$ in Supplementar y Equation (32) reduces to $e^{\alpha\langle n\rangle}$, since $e^x=\sum_{n=0}^{\infty}\frac{x^n}{n!}$.

Since $\langle n\rangle$ is not a directly measurable quantity in our experiment, we need a relation of $\langle n\rangle$ with experimental observables, e.g., protein concentration. Intuitively, $\langle n\rangle$ should bec ome larger when the free protein concentration $[P]_{FD}$ increases since in the absence of protein , $\langle n\rangle$ should be 0, and at very high concentration, $\langle n\rangle$ should approach n_o. And we found that at the low protein concentration limit, $\langle n\rangle$ is linearly proportional to $[P]_{FD}$, which is shown b elow.

The derivation of linear proportionality between $\langle n\rangle$ and $[P]_{FD}$ started from the rel ative population analysis between FD and TB states, which will be described in Supp lementary Section 1.5.7 below. In Supplementary Equation (44),

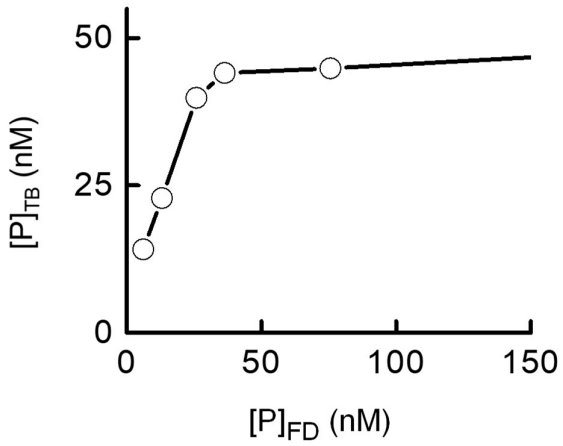
$\frac{[\mathrm{PD}]_{\mathrm{TB}}}{[\mathrm{P}]_{\mathrm{FD}}} = \frac{k_1[\mathrm{D}_0]_{\mathrm{TB}}}{k_{\mathrm{d}}} \frac{\partial \ln F_{\mathrm{TB}\leftarrow\mathrm{FD}}(x_{\mathrm{TB}\leftarrow\mathrm{FD}})}{\partial x_{\mathrm{TB}\leftarrow\mathrm{FD}}}$, here $[\mathrm{PD}]_{\mathrm{TB}}$, $[\mathrm{P}]_{\mathrm{FD}}$, and $[\mathrm{D}_0]_{\mathrm{TB}}$ are the cellular pr otein concentrations of TB and FD states, and the cellular concentration of TB sites, res pectively, and $F_{\mathrm{TB}\leftarrow\mathrm{FD}}(x_{\mathrm{TB}\leftarrow\mathrm{FD}}) \equiv \sum_{i=0}^{n_0} x_{\mathrm{TB}\leftarrow\mathrm{FD}}^{i}$, where $x_{\mathrm{TB}\leftarrow\mathrm{FD}} \equiv \frac{k_1}{k_{\mathrm{d}}}[\mathrm{P}]_{\mathrm{FD}}$. We used a low-conc entration-limit approximation where $F_{\mathrm{TB}\leftarrow\mathrm{FD}}(x_{\mathrm{TB}\leftarrow\mathrm{FD}}) \overset{[\mathrm{P}]_{\mathrm{FD}}\to 0}{\simeq} 1 + x_{\mathrm{TB}\leftarrow\mathrm{FD}} = 1 + \frac{k_1[\mathrm{P}]_{\mathrm{FD}}}{k_{\mathrm{o}}^{\mathrm{off}}}$ (note $k_{\mathrm{o}}^{\mathrm{off}}$ is the limiting value of k_{d} when $[\mathrm{P}]_{\mathrm{FD}}$ approaches zero and is also equivalent to the y-intercept of k_{d} -vs-$[\mathrm{P}]_{\mathrm{FD}}$ shown in Figure 2d; $k_{\mathrm{o}}^{\mathrm{off}}$ replaces k_{d} here under the approximation of low concentration; see also Supplementary Equation (35) later). Then we have:

$$\frac{[\mathrm{PD}]_{\mathrm{TB}}}{[\mathrm{P}]_{\mathrm{FD}}} = \frac{k_1[\mathrm{D}_0]_{\mathrm{TB}}}{k_{\mathrm{d}}} \frac{\partial \ln F_{\mathrm{TB}\leftarrow\mathrm{FD}}(x_{\mathrm{TB}\leftarrow\mathrm{FD}})}{\partial x_{\mathrm{TB}\leftarrow\mathrm{FD}}} \overset{[\mathrm{P}]_{\mathrm{FD}}\to 0}{\simeq} \frac{k_1[\mathrm{D}_0]_{\mathrm{TB}}}{k_{\mathrm{o}}^{\mathrm{off}}\left(1 + \frac{k_1[\mathrm{P}]_{\mathrm{FD}}}{k_{\mathrm{o}}^{\mathrm{off}}}\right)} \overset{[\mathrm{P}]_{\mathrm{FD}}\to 0}{\simeq} \frac{k_1[\mathrm{D}_0]_{\mathrm{TB}}}{k_{\mathrm{o}}^{\mathrm{off}}} \tag{33}$$

By definition, $\langle n \rangle = \frac{[\mathrm{PD}]_{\mathrm{TB}}}{[\mathrm{D}_0]_{\mathrm{TB}}}$, therefore by substituting Supplementary Equation (33) into $\langle n \rangle$, we can obtain the relation between $\langle n \rangle$ and $[\mathrm{P}]_{\mathrm{FD}}$.

$$\langle n \rangle = \frac{[\mathrm{PD}]_{\mathrm{TB}}}{[\mathrm{D}_0]_{\mathrm{TB}}} \overset{[\mathrm{P}]_{\mathrm{FD}}\to 0}{=} \frac{k_1}{k_{\mathrm{o}}^{\mathrm{off}}}[\mathrm{P}]_{\mathrm{FD}} \tag{34}$$

Supplementary Equation (34) predicts a linear relation between $[\mathrm{PD}]_{\mathrm{TB}}$ and $[\mathrm{P}]_{\mathrm{FD}}$ at the low protein concentration regime. This linear relation can be verified experimentally, as shown in Supplementary Figure 12.



Supplementary Figure 12. Dependence of $[P]_{TB}$ for Zur_{apo}^{mE} on $[P]_{FD}$. Linear dependence at the low free concentration regime (up to 50 nM) was observed, which verifies Supplementary Equation (34).

Then we plug in Supplementary Equation (34) into Supplementary Equation (32), which leads to:

$$k_{\mathrm{d}} = k_{\mathrm{o}}^{\mathrm{off}} + k_{\mathrm{r}}\left(e^{-\frac{[\mathrm{P}]_{\mathrm{FD}}}{K_{\mathrm{m}}}} - 1 \right) + k_{\mathrm{f}}[\mathrm{P}]_{\mathrm{FD}} \tag{35}$$

where $K_{\mathrm{m}} \equiv \frac{k_0^{\mathrm{off}}}{k_1(1-\alpha)}$, and $k_{\mathrm{o}}^{\mathrm{off}} \equiv k_{\mathrm{o}} + k_{\mathrm{r}}$,which represents a protein-concentration-independ ent unbinding rate constant. Supplementary Equation (35) gives the overall free protein c oncentration dependence of k_{d}. By fitting the experimentally measured unbinding rate co

nstant, we can obtain k_o^{off}, k_r, k_f, and K_m. Note that K_m has the concentration unit, and represent the effective affinity of oligomerization (that is, the lower the K_m, the stronger the oligomerization tendency). This equation shows that k_d approaches k_o^{off} when $[P]_{FD} \to 0$, which was used to derive supplementary equation (33). We also obtained $[P]_{FD}^{min}$, which is the free-protein concentration at which k_d reaches the minimum by $\frac{\partial k_d}{\partial [P]_{FD}}|_{[P]_{FD}=[P]_{FD}^{min}}=0$.

$$[P]_{FD}^{min} = K_m \ln\left(\frac{k_r}{k_f K_m}\right) \quad (36)$$

All fitted parameters are listed in Supplementary Table 6.

Overall mechanistic model of Zur-DNA interactions in cells

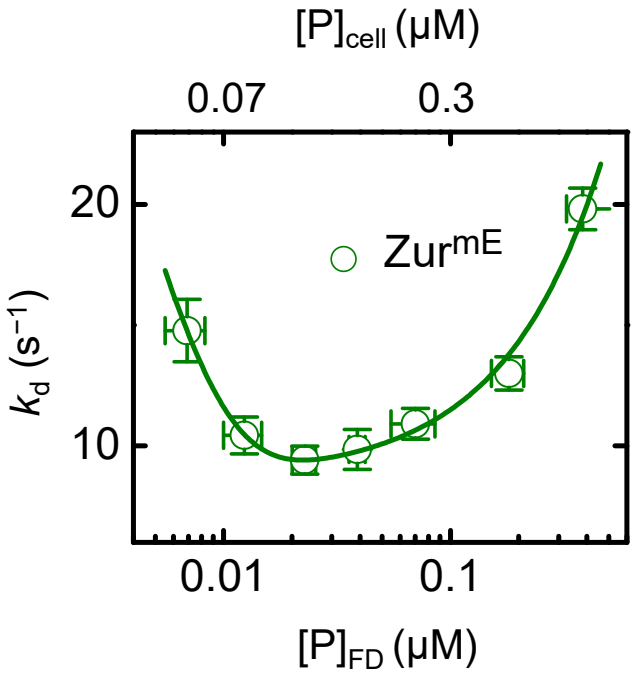
Based on the evidence of oligomerization of Zur on TB sites and three diffusion states comprising FD, NB, and TB states, we formulated a minimal mechanistic model for the overall Zur-DNA interaction in the cell (Figure 2c and Supplementary Figure 14 below).

Among the three states, this model has interconverting kinetic processes. Here, k_1 is a rate constant from FD to TB state, in which the apparent binding rate is linearly scaled with the effective concentration of vacant regulator TB sites $[D]_{TB}$ in the cell (that is, $k_1[D]_{TB}$). The apparent unbinding rate constant k_d includes three terms: the spontaneous, repressed, and facilitated unbinding as described in the previous section. The interconverting rate constants between FD and NB via binding nonspecifically to DNA are k_2 and k_{-2}, in which the apparent rate for binding is also linearly scaled with the effective concentration of vacant non-specific DNA binding sites $[D]_{NB}$ in the cell (that is, $k_2[D]_{NB}$). k_3 and k_{-3} are the interconversion rate constants between TB and NB. They are assumed to be very small because the number of TB

sites is very small relative to that of NB sites, which means when a regulator is at a NB site, TB sites are far away making the direct transition less probable. All the kinetic processes can be terminated by the photobleaching/blinking process of mEos3.2 tag with the rate constant k_{bl}.

Additional k_d results for Zur^{mE}

For Zur^{mE} in cells grown in medium without Zn^{2+} stress where the cellular Zur should be a mixture of apo and holo forms, its k_d still shows the biphasic behavior (Supplementary Figure 13). But the overall behavior of its k_d resembles more of Zur^{mE}_{apo}, further supporting that under 20 μM Zn^{2+} stress, the behavior of Zur^{mE}_{Zn} is dominated by the holo repressor form of Zur.



Supplementary Figure 13. Dependence of k_d on protein concentration for Zur^{mE}. Bottom/top

axis refers to free/cellular protein concentration, respectively. Lines are fits with Eq. (2). All error bars are s.d.

1.5.7 Additional kinetic and thermodynamic parameters from the analysis of relative populations of FD, NB, and TB states

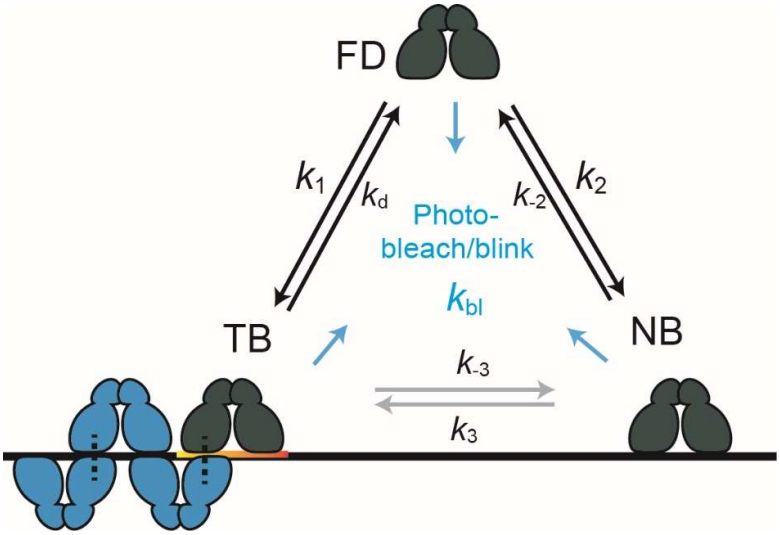
Analysis of relative populations of the FD, NB, and TB states using quasi-equilibrium model

In this section, we will introduce how to extract additional kinetic and thermodynamic quantities by analyzing the relative populations of the FD, NB, and TB states resolved by the analysis of the displacement distribution from single-molecule tracking of Zur^{mE} in the cell. The fastest diffusing state is assign as FD (i.e., freely diffusing Zur proteins in the cytosol), the medium diffusing state as NB (i.e., nonspecifically bound to DNA), and the slowest one as TB (i.e., tightly bound at a chromosomal site, either a consensus recognition site or any other tight-binding site that does not have a consensus sequence) in the cell.

Both kinetic and thermodynamic analyses are based on the kinetic model in Supplementary Figure 14, which includes the three states with the corresponding interconversion rate constants. Note that Zur protein can oligomerize at any TB site, with the number of Zur homodimers at any time at a site being n, which cannot be greater than n_o, the oligomerization number.

We assume that since the protein can sample the FD, NB, and TB states rapidly in the cell, we can approximate a quasi-equilibrium of binding and unbinding of regulators on DNA.

This approximation is valid when the time scale of interconversion between states (~ms, see Supplementary Section 0 and Supplementary Table 6) are much faster than the experimental imaging time (~30 min to 1 hour for each cell), during which the cellular Zur protein concentration does not vary much.



Supplementary Figure 14. Overall mechanistic model of Zur-DNA interactions in a cell. Interconversion rate constants are denoted *k*'s. Figure here is the same as Figure 2c in the main text.

Determination of binding rate constant (i.e., k_1) and affinity (i.e., $K_{D1} \equiv k_o^{off}/k_1$) to tight binding sites from relative populations of TB and FD states, $[PD]_{TB}/[P]_{FD}$

In this section, we will derive the equation to extract the binding rate constant (k_1) and affinity (K_{D1}) to the tight binding sites from relative populations of FD and TB states, focusing

on the quasi-equilibrium between the two states. The sequence of derivation is analogous (but with modifications) to that of Brunauer-Emmett-Teller (BET)[48] adsorption isotherm model for multilayer adsorption of molecules onto surface. Mechanistically, modeling of our protein-DNA interaction as a BET adsorption process is reasonable because the model explicitly describes sequential binding and unbinding of multiple proteins. The differences between the BET model and our model are that: 1) BET model assumes infinite number of layers for molecular adsorption, whereas our model has an upper limit being the oligomerization number (n_o) at each TB site. 2) BET model has two types of binding rate constant (to form the first layer and the rest), whereas our model has one binding rate constant. As the derivation of BET model did, one can define $\theta_0, \theta_1, \theta_2, \cdots, \theta_n, \cdots, \theta_{n_0}$, where n_o is the maximum number of Zur dimers at each TB site.

$$\theta_n = \frac{\text{The number of TB sites with } n \text{ dimers adsorbed}}{\text{Total number of TB sites}} \tag{37}$$

Note that the value of θ_n can vary from 0 to 1. At quasi-equilibrium, $\frac{d\theta_n}{dt} = 0$. With binding rate constant and unbinding rate constant, k_1 and k_d, respectively, we can obtain the expression of θ_n as a function of k_1 and k_d, which is shown below.

$$0 = \frac{d\theta_0}{dt} = k_d\theta_1 - k_1\theta_0[\mathrm{P}]_{FD}, \quad \theta_1 = \frac{k_1}{k_d}[\mathrm{P}]_{FD}\,\theta_0 \tag{38}$$

$$0 = \frac{d\theta_1}{dt} = -k_d\theta_1 + k_1\theta_0[\mathrm{P}]_{FD} + k_d\theta_2 - k_1\theta_1[\mathrm{P}]_{FD}, \quad \theta_2 = \frac{k_1}{k_d}[\mathrm{P}]_{FD}\,\theta_1 = \left(\frac{k_1}{k_d}[\mathrm{P}]_{FD}\right)^2\theta_0 \tag{39}$$

$$\therefore\ \ \theta_n = \left(\frac{k_1}{k_\mathrm{d}}[\mathrm{P}]_\mathrm{FD}\right)^n \theta_0 = x^n\theta_0,\ \ \left(x_{\mathrm{TB}\leftarrow\mathrm{FD}} \equiv \frac{k_1}{k_\mathrm{d}}[\mathrm{P}]_\mathrm{FD}, 0 \le n \le n_0\right) \tag{40}$$

Here, the unbinding rate constant k_d is a sum of the spontaneous unbinding rate constant, t he repressed unbinding term, and the facilitated unbinding term, and it depends on protei n concentration (i.e., $k_\mathrm{d} = k_\mathrm{o}^\mathrm{off} + k_\mathrm{r}\left(e^{-\frac{[P]_\mathrm{FD}}{K_\mathrm{m}}} - 1\right) + k_\mathrm{f}[\mathrm{P}]_\mathrm{FD}$,from Supplementary Equation (3 5) in Supplementary Section 0) and all other parameters are macroscopic quantities, whic h are discussed in Supplementary Section 0.

Note that, this model uses k_d, not $k_\mathrm{d}^{(n)}$, as an approximation. The microscopic unbin ding rate constant $k_\mathrm{d}^{(n)}$ depends on n, but assuming its behavior is ergodic and once it sa mples ergodically all possible values of n over a period of time [note our experimental i maging time (~30 min to 1 h) is orders of magnitude longer than the timescale of protei n binding/unbinding (~100 ms, see Supplementary Section 0), it should converge to k_d.

Then we can derive the average protein coverage at any TB site $\theta_{\mathrm{TB}\leftarrow\mathrm{FD}}$.

$$\begin{aligned}\theta_{\mathrm{TB}\leftarrow\mathrm{FD}} &= \sum_{i=0}^{n_0} i\theta_i = \frac{\sum_{i=0}^{n_0} i\theta_i}{1} \overset{\sum_{i=0}^{n_0}\theta_i=1}{=} \frac{\sum_{i=0}^{n_0} i\theta_i}{\sum_{i=0}^{n_0}\theta_i} = \frac{\theta_0\sum_{i=0}^{n_0} i x_{\mathrm{TB}\leftarrow\mathrm{FD}}^i}{\theta_0\sum_{i=0}^{n_0} x_{\mathrm{TB}\leftarrow\mathrm{FD}}^i} = \frac{x_{\mathrm{TB}\leftarrow\mathrm{FD}}\sum_{i=1}^{n_0} i x_{\mathrm{TB}\leftarrow\mathrm{FD}}^{i-1}}{\sum_{i=0}^{n_0} x_{\mathrm{TB}\leftarrow\mathrm{FD}}^i} \\ &= \frac{x_{\mathrm{TB}\leftarrow\mathrm{FD}}\left(1 + 2x_{\mathrm{TB}\leftarrow\mathrm{FD}} + 3x_{\mathrm{TB}\leftarrow\mathrm{FD}}^2 + \cdots + n_0 x_{\mathrm{TB}\leftarrow\mathrm{FD}}^{n_0-1}\right)}{\left(1 + x_{\mathrm{TB}\leftarrow\mathrm{FD}} + x_{\mathrm{TB}\leftarrow\mathrm{FD}}^2 + \cdots + x_{\mathrm{TB}\leftarrow\mathrm{FD}}^{n_0}\right)}\end{aligned} \tag{41}$$

$$\theta_{\mathrm{TB}\leftarrow\mathrm{FD}}=\frac{x_{\mathrm{TB}\leftarrow\mathrm{FD}}\dfrac{\partial F_{\mathrm{TB}\leftarrow\mathrm{FD}}\left(x_{\mathrm{TB}\leftarrow\mathrm{FD}}\right)}{\partial x_{\mathrm{TB}\leftarrow\mathrm{FD}}}}{F_{\mathrm{TB}\leftarrow\mathrm{FD}}\left(x_{\mathrm{TB}\leftarrow\mathrm{FD}}\right)}=x_{\mathrm{TB}\leftarrow\mathrm{FD}}\frac{\partial \ln F_{\mathrm{TB}\leftarrow\mathrm{FD}}\left(x_{\mathrm{TB}\leftarrow\mathrm{FD}}\right)}{\partial x_{\mathrm{TB}\leftarrow\mathrm{FD}}},\ \text{where} \tag{42}$$

$$F_{\mathrm{TB}\leftarrow\mathrm{FD}}\left(x_{\mathrm{TB}\leftarrow\mathrm{FD}}\right)\equiv 1+x_{\mathrm{TB}\leftarrow\mathrm{FD}}+x_{\mathrm{TB}\leftarrow\mathrm{FD}}^{2}+\cdots+x_{\mathrm{TB}\leftarrow\mathrm{FD}}^{n_0}=\sum_{i=0}^{n_0}x_{\mathrm{TB}\leftarrow\mathrm{FD}}^{i}$$

The value of $\theta_{\mathrm{TB}\leftarrow\mathrm{FD}}$ can vary from 0 to n_0. Since the total number (or concentration) of proteins bound at all TB sites equals the total number (or concentration) of TB sites m ultiplied by the average protein coverage at any TB site (i.e., $\theta_{\mathrm{TB}\leftarrow\mathrm{FD}}$,), one can write

$$\left[\mathrm{PD}\right]_{\mathrm{TB}}=\theta\left[\mathrm{D}_0\right]_{\mathrm{TB}}=\frac{k_1\left[\mathrm{P}\right]_{\mathrm{FD}}\left[\mathrm{D}_0\right]_{\mathrm{TB}}}{k_{\mathrm{d}}}\frac{\partial \ln F_{\mathrm{TB}\leftarrow\mathrm{FD}}\left(x_{\mathrm{TB}\leftarrow\mathrm{FD}}\right)}{\partial x_{\mathrm{TB}\leftarrow\mathrm{FD}}} \tag{43}$$

$$\frac{\left[\mathrm{PD}\right]_{\mathrm{TB}}}{\left[\mathrm{P}\right]_{\mathrm{FD}}}=\frac{k_1\left[\mathrm{D}_0\right]_{\mathrm{TB}}}{k_{\mathrm{d}}}\frac{\partial \ln F_{\mathrm{TB}\leftarrow\mathrm{FD}}\left(x_{\mathrm{TB}\leftarrow\mathrm{FD}}\right)}{\partial x_{\mathrm{TB}\leftarrow\mathrm{FD}}} \tag{44}$$

Supplementary Equation (44) shows how the ratio of $[\mathrm{PD}]_{\mathrm{TB}}$ and $[\mathrm{P}]_{\mathrm{FD}}$ relate to k_1 and k_{d}. By fitting $[\mathrm{PD}]_{\mathrm{TB}}/[\mathrm{P}]_{\mathrm{FD}}$ *vs* $[\mathrm{P}]_{\mathrm{FD}}$ with Supplementary Equation (44), we can extract k_1 and $K_{\mathrm{D1}}(\equiv k_{\mathrm{o}}^{\mathrm{off}}/k_1)$

Determination of K_{D3} ($=k_{-3}/k_3$) and $[D_0]_{TB}$ from $[PD]_{TB}/[PD]_{NB}$.

In this section, we will derive the connection between K_{D3} ($\equiv k_{-3}/k_3$), $[\mathrm{D}_0]_{\mathrm{TB}}$, and the relative populations of NB and TB states. The derivation of the equation for the ratio of $[\mathrm{PD}]_{\mathrm{TB}}$ and $[\mathrm{PD}]_{\mathrm{NB}}$ is analogous to that for the ratio of $[\mathrm{PD}]_{\mathrm{TB}}$ and $[\mathrm{P}]_{\mathrm{FD}}$, but the quasi-equilibrium under consideration is between the TB and NB states. The effective rate from TB to NB is $k_{-3}[\mathrm{D}]_{\mathrm{NB}}$, where $[\mathrm{D}]_{\mathrm{NB}}$ is the effective concentrations of vacant nonspecific DNA site. Since

$[D]_{NB} = [D_0]_{NB}-[PD]_{NB}$ where $[D_0]_{NB}$ is the effective concentration of total nonspecific binding sites in the cell, we can obtain the average protein coverage at any TB site that is in equilibrium with the NB state, $\theta_{TB\leftarrow NB}$

$$\theta_{TB\leftarrow NB} = \sum_{i=0}^{n_0} i\theta_{i,TB\leftarrow NB} = \frac{\sum_{i=0}^{n_0} i\theta_{i,TB\leftarrow NB}}{1} \overset{\sum_{i=0}^{n_0}\theta_{i,TB\leftarrow NB}=1}{=} \frac{\sum_{i=0}^{n_0} i\theta_{i,TB\leftarrow NB}}{\sum_{i=0}^{n_0}\theta_{i,TB\leftarrow NB}} = \frac{\theta_{0,NB}\sum_{i=0}^{n_0} i x_{TB\leftarrow NB}^{i}}{\theta_{0,NB}\sum_{i=0}^{n_0} x_{TB\leftarrow NB}^{i}} = \frac{x_{NB}\sum_{i=1}^{n_0} i x_{TB\leftarrow NB}^{i-}}{\sum_{i=0}^{n_0} x_{TB\leftarrow NB}^{i}}$$

$$= \frac{x_{TB\leftarrow NB}\left(1+2x_{TB\leftarrow NB}+3x_{TB\leftarrow NB}^{2}+\cdots+n_0 x_{TB\leftarrow NB}^{n_0-1}\right)}{\left(1+x_{TB\leftarrow NB}+x_{TB\leftarrow NB}^{2}+\cdots+x_{TB\leftarrow NB}^{n_0}\right)} \tag{45}$$

Here, $x_{TB\leftarrow NB} \equiv \frac{k_3}{k_{-3}\left([D_0]_{NB}-[PD]_{NB}\right)}[PD]_{NB}$, analogous to that in Supplementary Equation (40) earlier. Let $F_{TB\leftarrow NB}(x) \equiv 1+x_{TB\leftarrow NB}+x_{TB\leftarrow NB}^{2}+\cdots+x_{TB\leftarrow NB}^{n_0} = \sum_{i=0}^{n_0} x_{TB\leftarrow NB}^{i}$, then we hav e

$$\theta_{TB\leftarrow NB} = \frac{x_{TB\leftarrow NB}\dfrac{\partial F_{TB\leftarrow NB}\left(x_{TB\leftarrow NB}\right)}{\partial x_{TB\leftarrow NB}}}{F_{TB\leftarrow NB}\left(x_{TB\leftarrow NB}\right)} = x_{NB}\frac{\partial \ln F_{TB\leftarrow NB}\left(x_{TB\leftarrow NB}\right)}{\partial x_{TB\leftarrow NB}} \tag{46}$$

Since the total number of proteins bound at all TB sites equals to the product of the total number of TB sites and the average protein coverage at any TB site (i.e., $\theta_{TB\leftarrow NB}$), we can write:

$$[\mathrm{PD}]_{\mathrm{TB}} = \theta_{\mathrm{TB}\leftarrow\mathrm{NB}}[\mathrm{D}_0]_{\mathrm{TB}} = \frac{k_3[\mathrm{PD}]_{\mathrm{NB}}[\mathrm{D}_0]_{\mathrm{TB}}}{k_{-3}\left([\mathrm{D}_0]_{\mathrm{NB}} - [\mathrm{PD}]_{\mathrm{NB}}\right)} \frac{\partial \ln F_{\mathrm{TB}\leftarrow\mathrm{NB}}\left(x_{\mathrm{TB}\leftarrow\mathrm{NB}}\right)}{\partial x_{\mathrm{TB}\leftarrow\mathrm{NB}}} \quad (47)$$

$$\frac{[\mathrm{PD}]_{\mathrm{TB}}}{[\mathrm{PD}]_{\mathrm{NB}}} = \frac{k_3[\mathrm{D}_0]_{\mathrm{TB}}}{k_{-3}\left([\mathrm{D}_0]_{\mathrm{NB}} - [\mathrm{PD}]_{\mathrm{NB}}\right)} \frac{\partial \ln F_{\mathrm{TB}\leftarrow\mathrm{NB}}\left(x_{\mathrm{TB}\leftarrow\mathrm{NB}}\right)}{\partial x_{\mathrm{TB}\leftarrow\mathrm{NB}}} \quad (48)$$

Supplementary Equation (48) shows how the ratio of $[PD]_{TB}$ and $[PD]_{NB}$ relates to K_{D3} ($\equiv k_{-3}/k_3$) and $[D_0]_{TB}$. By fitting $[PD]_{TB}/[PD]_{NB}$ *vs* $[PD]_{NB}$ with Supplementary Equation (48), one can extract K_{D3} ($\equiv k_{-3}/k_3$) and $[D_0]_{TB}$.

Determination of K_{D2} ($\equiv k_{-2}/k_2$) and $[D_0]_{NB}$ from $[PD]_{NB}/[P]_{FD}$

Based on the same quasi-equilibrium approximation between the FD and NB states, the following relation holds:

$$\frac{[\mathrm{PD}]_{\mathrm{NB}}}{[\mathrm{P}]_{\mathrm{FD}}} = \frac{k_2[\mathrm{D}]_{\mathrm{NB}}}{k_{-2}} \quad (49)$$

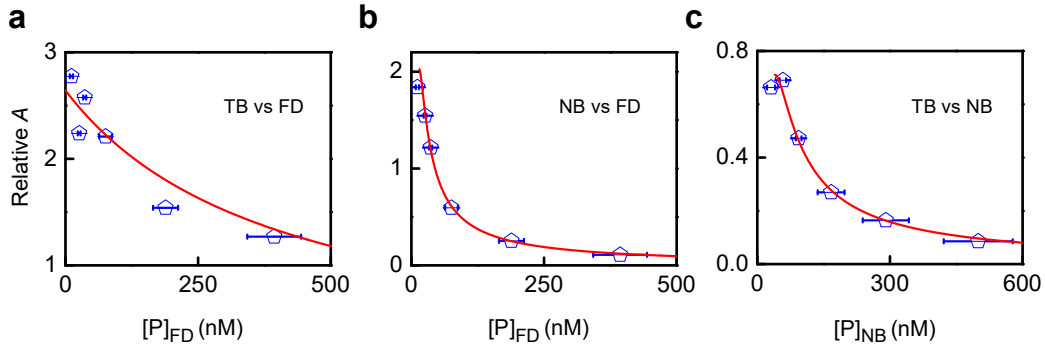
By substituting $[D]_{NB}$ with $[D_0]_{NB}$-$[PD]_{NB}$ in Supplementary Equation (49) and rearranging the equation leads to the final formula for K_{D2} ($=k_{-2}/k_2$) and $[D_0]_{NB}$.

$$\frac{[\mathrm{PD}]_{\mathrm{NB}}}{[\mathrm{P}]_{\mathrm{FD}}} = \frac{[\mathrm{D}_0]_{\mathrm{NB}}}{K_{\mathrm{D2}} + [\mathrm{P}]_{\mathrm{FD}}} \quad (50)$$

Supplementary Equation (50) shows the relation between $[PD]_{NB}/[P]_{]FD}$, K_{D2} ($\equiv k_{-2}/k_2$), and $[D_0]_{NB}$. By fitting $[PD]_{NB}/[P]_{FD}$ *vs* $[PD]_{NB}$ with Supplementary Equation (50), one can extract K_{D2} ($\equiv k_{-2}/k_2$) and $[D_0]_{NB}$.

Example of above population analysis

From the experimentally measured cellular protein concentration $[P]_{cell}$, we could obtain $[P]_{FD}$, $[PD]_{NB}$, $[PD]_{TB}$ via $[P]_{cell}\times A_{FD}$, $[P]_{cell}\times A_{NB}$, and $[P]_{cell}\times A_{TB}$, respectively, where A_{FD}, A_{NB}, and A_{TB} are the fractional populations of each diffusion state determined from the CDF analysis in Supplementary Section 1.5.4. We could then obtain additional kinetic parameters (e.g., k_1 and k_2) and thermodynamic parameters (e.g., K_{D1} and K_{D2}) from $[PD]_{TB}/[P]_{FD}$ vs. $[P]_{FD}$, $[PD]_{NB}/[P]_{FD}$ vs. $[P]_{FD}$, and $[PD]_{TB}/[P]_{NB}$ vs. $[PD]_{NB}$ (Supplementary equations (44), (48), and (50)) using values of k_o^{off}, k_r, k_f, and K_m pre-determined from analysis of k_d (e.g., Figure 2d). In the fitting routine, we used a fixed value (i.e., 5) for the oligomerization number (n_o), which is the maximum number of proteins at a TB site and should be greater than two (e.g., in the case of dimer of dimer is n_o=2). The dependence of the extracted parameters on the value of n_0 will be discussed in Supplementary Section 0.



Supplementary Figure 15. Example of relative population analysis of different diffusion states of Zur_{apo}^{mE} in cells. The fractional population data of (**a**) $[PD]_{TB}/[P]_{FD}$ vs. $[P]_{FD}$, (**b**) $[PD]_{NB}/[P]_{FD}$ vs. $[P]_{FD}$, and (**c**) $[PD]_{TB}/[P]_{NB}$ vs. $[PD]_{NB}$ are plotted as blue open circles. Red curves represent the fits with corresponding equations. Error bars are s.d.

Summary of extracted kinetic and thermodynamic parameters

We extracted the kinetic and thermodynamic parameters for all Zur^{mE} variants and summarized in Supplementary Table 6.

Supplementary Table 6. Kinetic and thermodynamic parameters for Zur-DNA interaction in *E.coli* cells

Parameters	Zur^{mE}	Zur^{mE}_{apo}	Zur^{mE}_{Zn} [b]	$Zur^{mE}_{apo,\ D49A}$	$Zur^{mE}_{Zn,\ D49A}$ [b]
k_1 ($nM^{-1}\ s^{-1}$) [a]	1.90 ± 0.17	1.84 ± 0.20	1.10 ± 0.18	1.61 ± 0.58	1.30 ± 0.19
k_o^{off} (s^{-1})	25 ± 12	22 ± 21	5.4 ± 0.6	22.1 ± 1.5	36 ± 41
k_r (s^{-1})	16 ± 11	12 ± 20	no. obs.[c]	20.8 ± 1.3	27 ± 40
k_f ($nM^{-1}\ s^{-1}$)	0.028 ± 0.005	0.044 ± 0.007	0.026 ± 0.033	0.049 ± 0.014	0.062 ± 0.010
K_m (nM)	6.0 ± 4.0	4.9 ± 7.3	no. obs. [c]	16.2 ± 7.5	3.2 ± 1.9
$[P]^{min}_{FD}$ (nM)	27.3 ± 3.5	19.8 ± 6.7	< 3.7	52.8 ± 3.4	15.6 ± 2.3
K_{d1} ($=k_o^{off}/\ k_1$) (nM) [a]	12.9 ± 6.2	11.7 ± 11.2	4.9 ± 1.2	13.7 ± 5.0	28 ± 20
K_{d2} ($=k_{-2}/\ k_2$) (nM) [a]	417 ± 35	348 ± 84	534 ± 148	209 ± 69	532 ± 134
K_{d3} ($=k_{-3}/\ k_3$) [a]	0.011 ± 0.002	0.023 ± 0.007	0.022 ± 0.023	0.032 ± 0.062	0.008 ± 0.006
$[D_0]_{TB}$(nM) [a]	8.51 ± 0.19	6.87 ± 0.65	10.9 ± 2.9	6.3 ± 1.0	7.75 ± 0.76
$[D_0]_{NB}$ (nM) [a]	1144 ± 84	961 ± 205	1201 ± 287	858 ± 230	1538 ± 353
$[D_0]_{TB} \cdot n_o$(nM) [a]	42.56 ± 0.94	34.3 ± 3.2	54 ± 14	31.6 ± 5.1	38.8 ± 3.8

[a] n_o = 5 was used in fitting.

[b] 20 μM of Zn^{2+} was used for zinc stress condition.

[c] Not observed.

Based on the extracted kinetic and thermodynamic parameters, we observed that:

Apo-Zur can bind to chromosome tightly (almost as tight as holo-Zur). The binding affinity of apo-protein (K_{d1} = 12.9 ± 6.2 nM) to TB sites is comparable to that of holo-protein (K_{d1}= 4.9 ± 1.2 nM), and much stronger than that to NB sites (K_{d2} = 417 ± 35 nM). The strong binding affinity of apo-Zur likely comes from its binding to non-consensus sites because previous study showed that apo-Zur does not bind to consensus operator sites[15].

The chromosome has similar binding capability for apo- and holo-Zur. The value of $[D_0]_{TB} \cdot n_o$ for all Zur variants ranges from 35 to 55 nM. Assuming $n_0 \geq 2$ for the oligomerization number, the concentration of Zur binding sites in the cell would be <25 nM, corresponding to <25 sites.

Salt-bridge mutation (D49A) weakens the oligomerization tendency. The K_m of holo-Zur was not measurable since it is lower than our experimentally accessible cellular protein concentration (~ 3.7 nM), whereas that of $Zur^{mE}_{Zn,\ D49A}$ became measurable and its value is 3.2 ± 1.9 nM. The K_m of apo-Zur is 4.9 ± 7.3 nM whereas that of $Zur^{mE}_{apo,\ D49A}$ is 16.2 ± 7.5 nM, reflecting a 2~3 times weakened oligomerization tendency.

Binding rate constants were observed to be similar among all Zur variants while the

unbinding rate constants varied significantly. Therefore, regulation of unbinding kinetics, rather than the binding kinetics, is a more critical factor in determining the binding affinity. For example, the binding rate constant k_1 to a TB site for Zur^{mE}_{apo} is 1.85 ± 0.19 nM^{-1} s^{-1}, slightly larger than that of Zur^{mE}_{Zn} (1.09 ± 0.18 nM^{-1} s^{-1}), indicating that apo-Zur's weaker binding to chromosome than holo-Zur is more dictated by their unbinding instead of binding rate constants.

Weakened oligomerization (i.e., larger K_m) leads to faster facilitated unbinding (i.e., larger k_f). The k_f for salt-bridge mutants of holo- and apo-Zur (0.062 ± 0.010 s^{-1}, and 0.049 ± 0.014 s^{-1}, respectively) are greater than that for non- mutated forms of holo- and apo-Zur (0.026 ± 0.033 s^{-1}, and 0.044 ± 0.007 s^{-1}, respectively). This is perhaps due to easier formation of the ternary complex where two Zur dimers each bind partly to DNA (Figure 2e, lower).

The parameters extracted for Zur^{mE} in the absence of zinc stress are between those of Zur^{mE}_{apo} and Zur^{mE}_{Zn}, consistent with its behavior being a mixture of apo- and holo-forms of Zur.

Validation of quasi-static system approximation for using a linear-combination of three-diffusion states

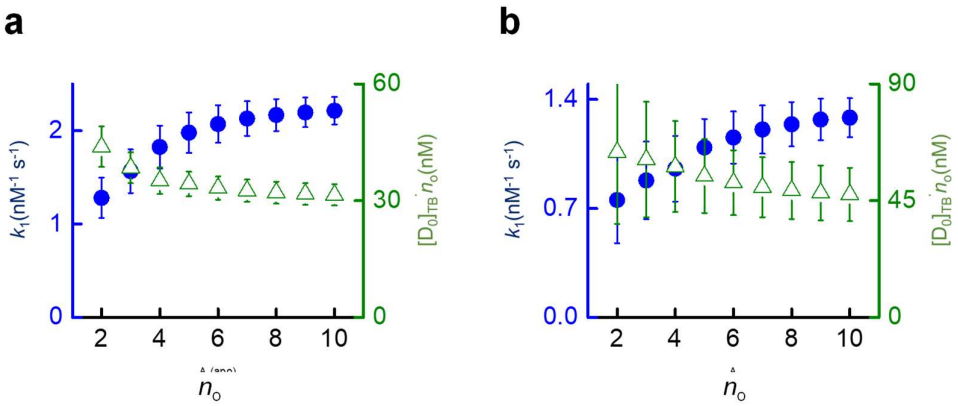
With extracted kinetic parameters, we can validate quasi-static system approximation in which we introduced in Section 0 for using a linear-combination of CDFs. Recall, to linearly combine three diffusion states of CDF, time-resolution in experiment should be comparable or faster than the interconversion rates between states. Using Zur^{mE}_{Zn} as an example, at the highest cellular protein concentration $[Zur^{mE}_{Zn}]_{cell}$ of ~940 nM (corresponding to $[Zur^{mE}_{Zn}]_{FD}$ ~410 nM) where all binding/unbinding steps are fastest, we examined the apparent interconversion rates between the states and their corresponding time scales. The computed unbinding rates of TB

(k_d) and NB (k_{-2}) are 17.2 and 12.1 s^{-1} (corresponding timescales are ~58 and 83 ms, respectively). The apparent binding rates of TB ($k_1[D]_{TB}$) and NB ($k_2[D]_{NB}$) are ~1.39 and 12.33 s^{-1} (corresponding timescales are 722 and 81 ms, respectively). Imaging time lapse of 40 ms is faster than these timescales of interconversion between states, which can validate the quasi-static system approximation.

n_o dependence of extracted kinetic and thermodynamic parameters

Previous *in vitro*, and *in silico* studies identified four *zur*-regulons including zinc uptake gene cluster (*znuABC*), ribosomal protein (*L31p* and *L36p*), periplasmic zinc trafficking protein (*zinT*), and lysozyme inhibitor (*pliG*) [12-15]. These 4 known consensus sites essentially set the lower limit for $[D_0]_{TB}$ for holo-Zur (i.e., ~4 nM; note 1 copy corresponds to 1 nM in a *E. coli* cell). We obtained 35 - 55 nM for $[D_0]_{TB} \cdot n_o$ for all Zur variants indicating that n_o is <10. In our fitting routine, we had to choose a value of n_0 in advance, required by our numerical fitting algorithm, and we chose to use a fixed oligomerization number $n_o = 5$. Here we evaluate the effect of the value of n_0 on the fitting results. Two parameters, k_1 and $[D_0]_{TB} \cdot n_o$ were chosen to show the n_0 dependence (note, k_1 is the parameter in the summation; $F_{TB \leftarrow FD}\left([P]_{FD}\right) = \sum_{i=0}^{n_0}\left(\frac{k_1}{k_d}[P]_{FD}\right)^i$ in Supplementary Equation (44), meaning that this parameter is directly influenced by n_o in fitting). If these two asymptotically approach certain values, other thermodynamic quantities such as K_{d2} and $[D_0]_{NB}$ should approach their asymptotic values since parameters in fitting are all shared (see supplementary equation (44), (48) and (50). Supplementary Figure 16 shows that in the range where n_o >3, the extracted k_1 and $[D_0]_{TB} \cdot n_o$ approach to asymptotic values within their error bars. Therefore, the

oligomerization number n_o does not significantly influence the extracted parameters.



Supplementary Figure 16. n_o dependence of the extracted k_1 (blue circles) and $[D_0]_{TB} \cdot n_o$ (green triangles) for (**a**) Zur^{mE}_{apo} and (b) Zur^{mE}_{Zn}. Error bars are s.d.

1.5.8 Clustering analysis of residence sites

Pairwise distance distributions (PWD) analysis of residence sites and the simulated random sites

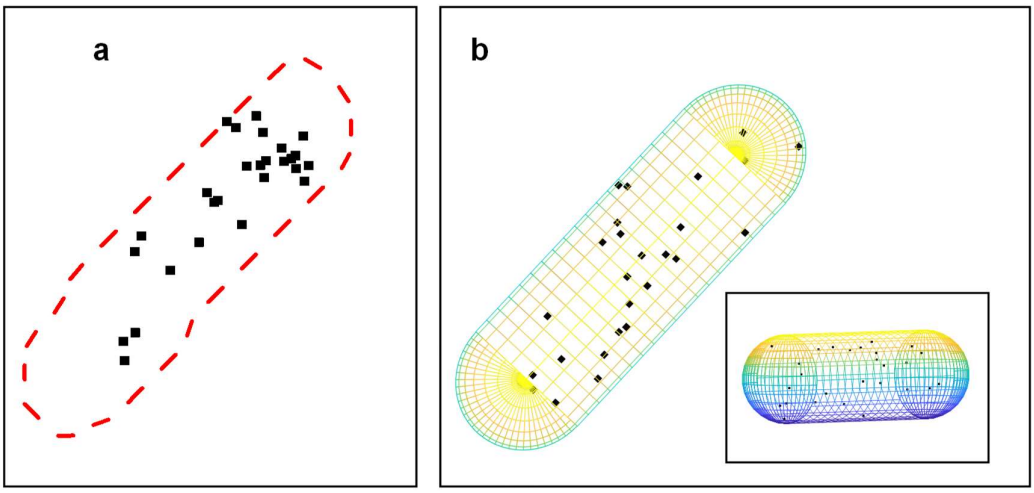
Residence sites are the locations associated with the residence times of Zur^{mE} obtained by r_o-thresholding the displacement vs. time trajectories (Supplementary Section 1.5.5), which are dominated by Zur binding to the tight binding sites on the chromosome. An example of residence sites in a cell are plotted in Supplementary Figure 17a with the corresponding cell outline, which was determined from the cell's optical transmission image.

Zur^{mE} clustering was first examined by computing the pairwise distance distribution

(PWD) of the residence sites. The pairwise distance is the Euclidean distance between positions (i.e., sites). Only the 1st position within each residence time was sampled in computing the PWD to avoid oversampling toward relatively longer residence times. Residence sites in each group of cells having a similar protein concentration were compiled to minimize skewing the PWD toward cells of higher cellular protein concentrations that have higher densities of sampled points. Within each group of cells of a similar protein concentration, those cells having more clustered residence sites are expected to have shorter pair-wise distances.

As a control, identical numbers of randomized locations were simulated using a home-written MATLAB program within each model cell geometry (i.e., a cylinder with 2 hemisphere caps) whose length and width were taken as the averages of the cells in each group, measured from the optical transmission images (Supplementary Section 1.5.3). Sampling was repeated thirty times and averaged for sufficient statistics. The simulation was also done in 3-dimension, and a subsequent 2-D projection was used to directly compare with the 2-D experimental data (Supplementary Figure 17b).

The PWDs of residence sites for both the Zur^{mE}_{apo} and Zur^{mE}_{Zn} show a peak at a shorter distance than that from the simulation of randomly distributed locations, implying the clustering of Zur^{mE} proteins in the cell, presumably due to oligomerization at tight binding sites on the chromosome (Figure 3a and Supplementary Figure 18a).



Supplementary Figure 17. (**a**) An example of residence sites (black squares) of Zur_{apo}^{mE} in a cell. The cell outline is indicated by the dashed red line. (**b**) *xy*-projected view of randomly sampled 3-D points within a model cell of the same length and width. Inset: Same positions with a *yz*-projected view.

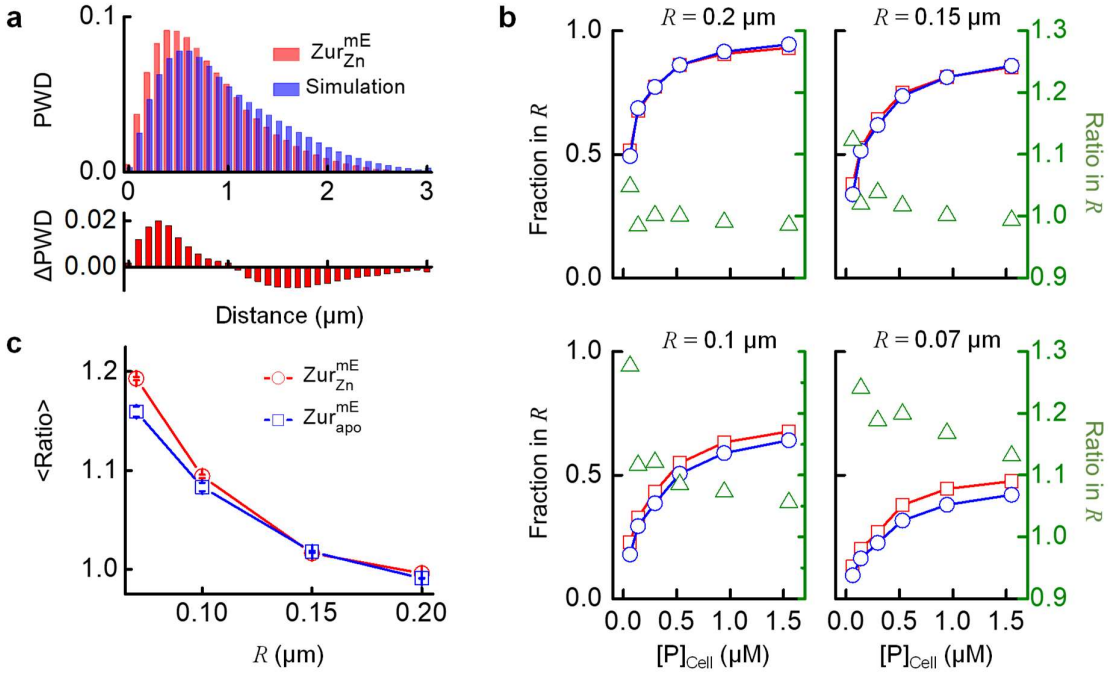
Analysis of the fraction of residence sites within a threshold distance R to decouple the effect of chromosome condensation

Our previous research on the metalloregulator CueR and ZntR used the averaged pairwise distance between residence sites, $\langle d_{ij} \rangle$ as a measure for the extent of chromosome compactness, where cells with more compacted chromosome have shorter $\langle d_{ij} \rangle$;[11] that analysis was based on that CueR/ZntR has a large number of recognition sites scattered across the chromosome randomly and that CueR/ZntR do not show any oligomerization behaviors. The quantified number of $\langle d_{ij} \rangle$ was ~0.68 μm for compacted chromosome, and 0.88 μm for less

compacted chromosome. Since the extent of chromosome compactness differs from cell to cell, the $\langle d_{ij} \rangle$ are broadly distributed among cells implying that chromosome compactness (i.e., condensation) could contribute to shorter distances in the pair-wise distance distributions of residence sites.

To decouple the contribution of protein oligomerization from chromosome condensation in the PWD analysis, we computed the fraction of residence sites within a radius threshold distance R and compared with that of simulated random sites. Figure 3b and Supplementary Figure 18b present the fraction of residence sites with different values of R as a function of cellular protein concentration. For both the experimental and simulated residence sites, the fraction of residence sites within R increases with increasing protein concentration, because higher concentrations gave higher densities of sampled residence sites in a cell. When R became shorter from 0.2 μm down to 0.07 μm, the increase was slower (Supplementary Figure 18).

On the other hand, the difference of the fraction within R between the experimental data and simulated random sites becomes larger as R is smaller (Supplementary Figure 18b, red vs. blue). We calculated the averaged ratio within R of the fraction of observed residence sites over that of the simulated random sites; this average ratio shows a clear increase with decreasing R (Supplementary Figure 18c). This trend supports the occurrence of oligomerization of Zur^{mE} at TB sites because protein oligomerization is a molecular scale process (i.e., within short distances) whereas chromosome compactness is much larger in dimension (i.e., around hundreds of nanometer), and at decreasing threshold distance R, protein oligomerization would dominate over chromosome condensation in producing clustering of residence sites.



Supplementary Figure 18. Clustering analysis of Zur's residence sites in cells. (**a**) Normalized PWD of residence sites for Zur^{mE}_{Zn}, and of simulated random localizations (top) at $[Zur^{mE}_{Zn}]$ = 538 ± 87 nM, and the difference of the normalized PWD of Zur^{mE}_{Zn} from that of simulation (bottom). (**b**) The fraction of residence sites within the threshold distance R (left axis) for Zur^{mE}_{Zn} (red curve) and simulated random locations (blue curve) vs cellular protein concentration, as well as their ratio (green triangle, right axis). Each panel represents a different R value. (**c**) Dependence of the averaged ratio in (**b**) on R for both Zur^{mE}_{Zn} and Zur^{mE}_{apo}.

CHAPTER 2

QUANTIFYING MULTISTATE CYTOPLASMIC MOLECULAR DIFFUSION IN BACTERIAL CELLS VIA INVERSE TRANSFORM OF CONFINED DISPLACEMENT DISTRIBUTION

2.1 Introduction

Diffusive behaviors of membrane and cytoplasmic molecules in cells carry valuable information on the underlying biological processes, such as membrane protein oligomerization[58], protein-membrane interactions[59], protein-DNA interactions[60], DNA repair[61], cytokinesis[62], and chromosome diffusion[63]. As these processes fulfill many cellular functions, quantifying the diffusive behaviors of these molecules is important for understanding the underlying mechanisms.

A number of techniques have been developed to study the diffusive behaviors of membrane and cytoplasmic molecules. Fluorescence recovery after photobleaching (FRAP)[64], fluorescence correlation spectroscopy (FCS)[65], and single-molecule tracking (SMT)[66] are the three most common fluorescence-based methods[67]. Both FRAP and FCS probe molecular diffusive behaviors within a small volume defined by the laser focus. However, the slow time resolution and potential DNA damage caused by photobleaching in FRAP[68], the susceptibility to optical aberrations in FCS[69], and the diffraction-limited spatial resolution constrains the application of FRAP and FCS to molecular diffusions in live cells. On the other hand, recent technological advances in camera, fluorescent protein (FP) reporters, and super-resolution

imaging algorithm[70] made it possible to track individual molecules with high spatial (few nm) and temporal (μs) resolution[71] in live cells[72]. Imaging one molecule at a time typically is through imaging a fluorescent tag, which is often a regular or photoconvertible FP. Even though the photobleaching of the fluorescent tag limits the observation time, recent studies have shown that SMT is particularly powerful in dissecting the mechanisms of biophysical processes[73, 74]. Using probes such as quantum dots or plasmonic nanoparticles can further extend SMT trajectories in time[75].

Through real-time SMT, one directly obtains the diffusive behavior of each fluorescently labeled protein molecule in the cell reflected by its location versus time trajectory. Quantitative methods to analyze the SMT trajectories include mean-squared displacement (MSD), hidden Markov modeling (HMM)[76-79], and probability distribution function (PDF) or cumulative distribution function (CDF) of displacement length analyses. MSD analysis, the most popular method, reliably determines the diffusion coefficient for molecules moving in free space with a single diffusion state[80]. For molecules having transient diffusive behaviors or those containing multiple diffusion states, MSD method is less ideal due to its requirement of averaging over all displacements[81]. HMM analysis, a probabilistic maximum-likelihood algorithm, can extract the number of diffusion states and their interconversion rate constants (with certain assumptions)[78, 79, 82]; it provides a mathematically derived routine and unbiasedly analyses SMT trajectories, but the resulting multi-state diffusion model often lacks a definitive number of states[83]. The HMM analysis of SMT trajectories is further constrained by the complex computational algorithm and the difficulty in incorporating the photophysical kinetics

of the fluorescent probe. Analysis of the PDF or CDF of displacement length on the basis of Brownian diffusion model is known as a robust way to quantify the diffusion coefficients and fractional populations of multi-state systems, as demonstrated both *in vitro* and *in vivo*[60-62, 84-86], even though it requires more control experiments and elaborate analysis based on a defined kinetic model in order to extract the minimal number of diffusion states and their interconversion rate constants.

One factor that significantly affects the PDF or CDF analysis of cytoplasmic diffusion displacement is the confinement by the cell volume, especially for bacterial cells, which are less than a few microns in size. This confinement distorts and compresses the displacement length distribution, especially for molecules with large diffusion coefficients. SMT trajectories obtained from cells with different geometries can give significantly biased displacement length distributions, even though the underlying diffusion coefficient is the same. As a result, fitting the distribution of displacement length with PDF or CDF derived from the Brownian diffusion model (or any other model) only reports apparent diffusion coefficients, which are typically smaller than the intrinsic diffusion coefficients.

For membrane protein diffusion, it is a two-dimensional (2D) diffusion on a surface curved in three-dimensional (3D) space, and it does not actually have boundary confinement, as the cell membrane is a continuous boundary-less surface. However, SMT trajectories are generally obtained in 2D, where only the x, y movements in the imaging plane are tracked, thus projecting the boundary-less movements of membrane protein diffusion into a 2D diffusion

confined by the cell boundary. This confinement effect from 2D projection of membrane diffusion distorts and compresses the displacement length distribution as well. To address this projection-induced confinement effect, Peterman and coworkers introduced the inverse projection of displacement distribution (IPODD) method[87] in analyzing simulated 1-state membrane diffusion in bacterial cells (e.g., *E. coli*). In short, they first created a projected displacement distribution (PDD) matrix for a given cell geometry by projecting the simulated membrane displacement vectors onto the 2D imaging plane. For *each* displacement length that could occur anywhere on the membrane surface, they determined the resulting distribution of displacement length after projection. The PDD matrix thus quantifies the relationship from the displacement distribution before projection to that after projection. Using inverse transformation, they could then convert the 2D-projected displacement length distribution (which is often the one determined experimentally) into a most probable displacement length distribution on the cell membrane, which is readily analyzed to give the intrinsic diffusion coefficient.

Here we report an extension of the inverse transformation method for membrane diffusion to analyze cytoplasmic molecular diffusions. Using simulated diffusion trajectories in free and confined spaces, we demonstrate this inverse transformation method in analyzing 1-state cytoplasmic Brownian diffusions in both 2D and 3D and with varying diffusion coefficients and cell geometries. We further extend this method to multi-state cytoplasmic diffusions, containing non-interconverting or interconverting states, to effectively extract the minimal number of diffusion states as well as their respective diffusion coefficients and

fractional populations. Finally, we demonstrate a successful application to experimental SMT data of a transcription factor in living *E. coli* cells, which shows interconverting multi-state diffusive behaviors.

2.2 Method

2.2.1 Simulations of single-molecule diffusion trajectories

Based on the Brownian diffusion model, we used home-written Matlab codes to simulate three-dimensional (3D) single-molecule diffusion trajectories that contained one, three non-interconverting, or three interconverting diffusion states in both free space and confined space. Each simulation condition contained at least 100,000 diffusion trajectories to ensure statistically saturated data for analysis. The two-dimensional (2D) diffusion trajectories were generated from the 3D ones by discarding the z-component.

Diffusion trajectories in free space. The 3D diffusion trajectories in free space containing one diffusion state were simulated via the following steps. First, randomly sampled the initial position (x, y, z) in free space, where the values of x, y, and z are each from a randomly generated number. Second, with the input diffusion coefficient D, generated the distribution of displacement vector ($\vec{r_i}$, where i = x, y, or z) following Brownian diffusion in free space as described by Eq (51), where n = 1, for each of the three dimensions (i.e., x, y, and z) and using a time resolution t = 4 or 60 ms. Third, randomly chose a $\vec{r_i}$ from the distribution of the displacement vector, together with the initial position, to calculate the subsequent position,

which also served as the new initial position for the next simulation step. The procedure was then repeated until the length of the final moving trajectory contained 10 positions for analysis. Trajectories for three non-interconverting states were generated as that in single diffusion state case but with D of 0.036, 0.7, and 11 $\mu m^2 s^{-1}$, separately.

The 3D diffusion trajectories that contained three interconverting diffusion states were simulated with three input diffusion coefficients D_i (i = 1, 2, or 3) and their associated interconversion rate constants (e.g., rate constant γ_{ij} for interconversion from state i to j; $i \neq j$ and i, j = 1, 2, or 3). A sequence of residence time on the diffusion state i was built, where each residence time t_i sampled the residence time distribution $\exp(-\Sigma_j \gamma_{ij} t)$, where Σ_j was a sum of all competing processes leaving from state i to state j ($j \neq i$), each with a rate constant γ_{ij}. The transition from state i to a particular state j followed the relative probability $\frac{\gamma_{ij}}{\sum_j \gamma_{ij}}$. The residence time sequence was terminated by t_{bl}, which equaled the sum of all residence times in the sequence, and t_{bl} samples the distribution $\exp(-k_{bl} \frac{T_{int}}{T_{tl}} t)$, which was limited by the photobleaching and photoblinking of the fluorescence tag where k_{bl} is the tag's intrinsic photobleaching and photoblinking rate constant, T_{int} and T_{tl} are the laser exposure time and stroboscopic imaging lapse time, respectively. During each state, the generation of displacements was the same as described in the one diffusion state case. Here we first generate the primary diffusion trajectories with T_{int} and T_{tl} of 4 ms. For trajectories with longer T_{tl}, the primary diffusion trajectories were re-simulated and sampled at every lapse time T_{tl} to give the

eventual simulated diffusion trajectory, which is analyzed.

Diffusion trajectories in confined space. To mimic the 3D SMT data in a bacterium cell, we first modelled the 3D cell geometry as a cylinder capped by two hemi-spheres for simplicity with cell length and width adapted from our experimental results. The 3D diffusion trajectories in a confined space (i.e., the cell volume) were generated by similar procedures as described in free space, but with random selection of initial positions inside the cell volume and the implementation of confinement effect with the boundary reflection from the cell surface. Boundary reflection was performed when the end point of displacement vector is outside the cell volume. The intercept of cell boundary and displacement vector, together with the normal plane, were calculated for subsequent evaluation of the reflected position. The corresponding 2D simulated data were then generated from 3D ones by discarding the diffusion information in z direction. The 3D diffusion trajectories for systems with three interconverting diffusion states in confined space were simulated in the same way as in free space but with applied boundary reflection in the displacement generation step.

2.2.2 Generation of confinement transformation matrix ([CTM])

Generation of the confinement transformation matrix ([CTM]) was inspired by Peterman's work on inverse projection of displacement distributions (IPODD) for analyzing membrane proteins diffusing on the curved surface. In short, > 100,000 displacement vectors ($\vec{r}$) of a given distance length *r* were randomly positioned in the cell. If the end point of

displacement vector was outside the cell volume, the boundary reflection was performed, generating final positions. We then calculated the output *r* from the final positions and created the confined displacement distribution (CDD), which served as a single column data for the [CTM]. The length of displacement vector varied from 10 nm to 2.82 μm (i.e., up to the cell length) with 10 nm and 30 nm increments for transforming simulated and experimental data, respectively. Finally, CDDs for all input displacement vectors were combined to form the confinement transformation matrix.

2.2.3 Generation of probability density function (PDF) of displacement length for systems with multi diffusion states

All probability density functions of displacement length (PDF(*r*)) in this study were generated from the distribution of displacement length of moving trajectories normalized by the area of distribution. For example, for systems with a single diffusion state, displacements were calculated from the moving trajectory and used to generate the histogram of displacement length for a given bin size (i.e., 10 and 30 nm for simulated and experimental data, respectively). The displacement histogram was then divided by its area to create the probability density function of displacement length, PDF(*r*).

PDF(*r*) for systems with static three diffusion states was obtained as follows. We combined displacements from the respective diffusion states with given weighting coefficients to generate the displacement length histogram, which was then normalized by its area to create

the PDF(r) for analysis. For example, after simulating 100,000 trajectories (with trajectory length of 10 positions) in a given cell geometry for each of three different D_{input}, we randomly chose trajectories from each diffusion state and combined them with chosen fractional populations for subsequent analyses.

Finally, for system with three interconverting diffusion states, the PDF(r) was generated from moving trajectories based on procedures as described in Method Section 2.2. Since the moving trajectories were simulated with three interconverting diffusion states built-in, the PDF(r) was simply the resulting displacement histogram normalized by the histogram area.

2.2.4 Transformation of distribution of displacement length between free and confined spaces

Transformation of distribution of displacement length between free and confined spaces was achieved via the confinement transformation matrix ([CTM]). Forward converting the 2D or 3D distribution of displacement length in free space to that in confined space was via direct multiplication of the 2D or 3D distribution in free space with [CTM]. As for the inverse transformation (i.e., distribution in confined space to that in free space) process, the inverted [CTM] (i.e., $[CTM]^{-1}$) was first obtained using Gaussian elimination; multiplication of the distribution of displacement length in confined space with the $[CTM]^{-1}$ (i.e., Eq (54)) then resulted in the corresponding distribution in free space. Note that in the $[CTM]^{-1}$ obtaining step,

we first diagonalized the [CTM] and back substituted the known variables to solve for $[CTM]^{-1}$ rather than simply transpose the [CTM].

2.3 Results and Discussions

2.3.1 Inverse transform of confined displacement distribution for cytoplasmic molecules

The diffusive motions of cytoplasmic molecules in a bacterial cell are significantly confined by the small cell size (Figure A, right) — a typical *E. coli* cell is about 0.5 × 0.5 × 2 μm^3 in size (e.g., ~1.5 fL), and for a small protein with a diffusion coefficient of 10 $\mu m^2 s^{-1}$, its diffusion can traverse the cell length in ~100 ms. This confinement effect distorts the molecule's displacement distribution, hindering the quantification of its diffusion coefficient. For heterogeneous diffusion where multiple diffusion states are present, this confinement effect also hinders the determination of the (minimal) number of diffusion states. Here we present an inverse transform method to analyze displacement distributions of confined diffusions to obtain displacement distributions that are well described by Brownian diffusion in free space. The feasibility of the method is examined by diffusion simulations in free and confined spaces.

For 3D Brownian diffusion in free space, the probability density distribution of displacement vector $\vec{r}$ within time t follows a Gaussian function:

$$P(\vec{r},t)=\left(\frac{1}{\sqrt{4\pi Dt}}\right)^n \exp\left(-\frac{\vec{r}^{\,2}}{4Dt}\right) \tag{51}$$

where D is the diffusion coefficient and n = 1, 2, or 3 for 1D, 2D or 3D diffusion, respectively. The second moment of $\vec{r}$ follows the well-known relationship $\left\langle \vec{r}^2 \right\rangle = 2nDt$. The probability distribution function (PDF) of the 3D and 2D displacement length r, which is the scalar component of $\vec{r}$, can be obtained by integrating $P(\vec{r},t)$ over all angular spaces:

$$\mathrm{PDF}(r,t) = \frac{r^2}{\sqrt{4\pi}(Dt)^{3/2}} \exp\left(-\frac{r^2}{4Dt}\right), \text{(3D)} \tag{52}$$

$$\mathrm{PDF}(r,t) = \frac{r}{2Dt} \exp\left(-\frac{r^2}{4Dt}\right), \text{(2D)} \tag{53}$$

The blue shade in Figure D shows the distribution of displacement length r from a simulation of 3D Brownian diffusion in free space with D = 1 $\mu m^2 s^{-1}$ and t = 60 ms (simulation details in Method Section 2.2), which is well described by Eq (52) (Supplementary Figure 19). As most of the SMT experiments are done in 2D imaging mode, the blue shade in Figure F presents the distribution of the corresponding displacement length r in 2D, which is again well described by Eq (53)) (Supplementary Figure 19). When the same Brownian diffusion is simulated inside a confined space (e.g., inside a bacterial cell, Figure A, right), the distributions of displacement length r in both 3D and 2D are significantly distorted due to reflections by cell boundaries (Figure D and F), as expected. These confined displacement length distributions do not follow Eq (52) and (53), and attempted fitting gives the diffusion constant of 0.76±0.32 $\mu m^2 s^{-1}$, underestimated from the expected diffusion coefficient of 1 $\mu m^2 s^{-1}$.

To numerically mimic the confinement effect on the displacement length distribution, we followed Peterman *et al*[87] to generate a confinement transformation matrix ([CTM]; e.g.,

Figure B) for a given cell geometry, which is readily measured for bacterial cells. For each column of this matrix, a 3D displacement vector in free space of a given length is randomly sampled within the cell volume and applied boundary reflections when the vector impinges on the cell boundary. In this way, it generates a distribution of corresponding 3D displacement in the confined space. Normalizing this distribution gives the confined displacement distribution (CDD), which represents the probability distribution for finding a 3D-confined displacement length given a 3D displacement of a particular length in free space (Figure C). Varying the length of the 3D displacement vector in free space and repeating the random sampling process generates the data for all other columns in [CTM] (Figure B). The utility of this confinement transformation matrix can be seen by applying it to the distribution of displacement length from the simulated 3D Brownian diffusion in free space as $[\mathrm{CTM}]\cdot \mathrm{PDF}_{\mathrm{FS}} = \mathrm{PDF}_{\mathrm{CS}}$, where $\mathrm{PDF}_{\mathrm{FS}}$ and $\mathrm{PDF}_{\mathrm{CS}}$ are probability distribution functions of displacement length in free and confined spaces, respectively. The resulting distribution from this forward transformation reproduces that from the simulations in the confined space (Supplementary Figure 20A).

More useful is the inverse transformation of the confined displacement distribution (ITCDD), as the CDD is what is directly measured in experiments:

$$\mathrm{PDF}_{\mathrm{FS}} = [\mathrm{CTM}]^{-1} \cdot \mathrm{PDF}_{\mathrm{CS}} \tag{54}$$

where $[\mathrm{CTM}]^{-1}$ can be obtained by Gaussian elimination (Method Section 2.2.4). Applying ITCDD on the simulated results in the confined space deconvolutes the confinement effect and reproduces effectively the theoretical distribution of displacement length r in free space (Figure D). Fitting the inverse transformed distribution gives D = 1.04 ± 0.01 $\mu m^2 s^{-1}$,

recovering reliably the expected diffusion coefficient (D = 1 $\mu m^2 s^{-1}$). All fittings of the ITCDD were done via least-square fitting in MATLAB program, using PDFs with single or three diffusion states (e.g., Eq (3) and Eq (5), respectively, for the 2D diffusion cases)

The forward transformation, and more importantly, the inverse transformation using the confinement transformation matrix are equally applicable between the 2D displacement length distribution in free space and that in confined space (Figure E-F and Supplementary Figure 20B).

It is important to point out that this forward and inverse transformations only work well when the confinement transformation matrix is generated when the input and output displacements match in dimension. Figure G shows the [CTM] generated between 3D displacement in free space and the 2D displacement in confined space. Using this [CTM] or $[CTM]^{-1}$ for forward or inverse transformation cannot reproduce the expected distributions (Figure H and Supplementary Figure 20C). It is worth noting that the original Peterman's work on membrane diffusion is between 2D diffusion in curved surface and its 2D projection onto a flat surface[87], where the displacement dimensions are matching. A likely reason for the inapplicability of transformation between different dimensions is that the lower dimension displacements are missing information about the 3rd dimension; this missed dimension cannot be created during the transformation to the higher dimension displacements.

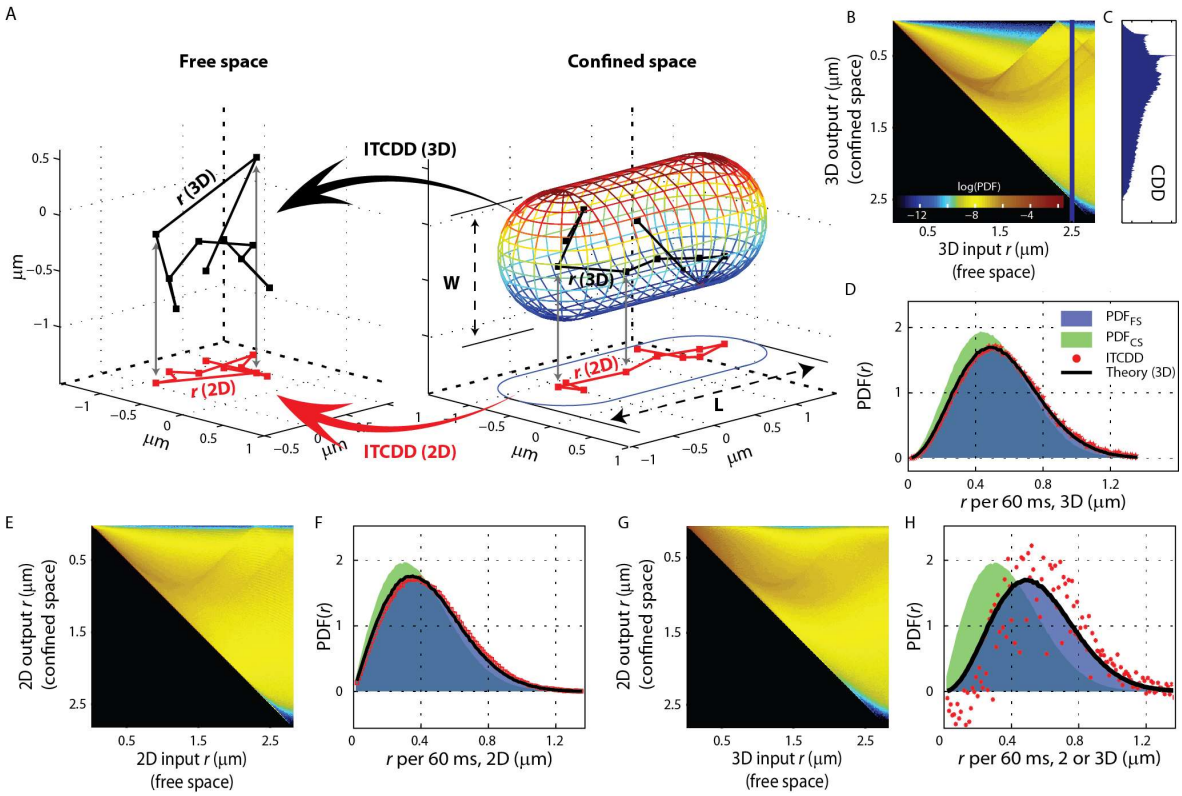


Figure 6. Illustration of inverse transform of confined displacement distribution (ITCDD) using simulated Brownian diffusions. (**A**) Schematic overview of ITCDD. Single-molecule diffusion trajectories are first generated in 3D in free or confined space (black trajectories) with t = 60 ms. Removing the z component from 3D trajectories results in the corresponding projected 2D trajectories (red trajectories). Converting the displacement length distribution in confined space to that in free space is achieved via inverse transformation of confined displacement distribution using the confinement transformation matrix ([CTM]). Here all confined diffusion simulations were performed within a cell having width (W) and length (L) of 1.15 and 2.82 μm respectively. (**B**) [CTM] for the 3D output displacements in confined space given 3D input displacements in free space. The input r is from 10 nm to 2820 nm with 10 nm increment. (**C**) CDD from B at 3D input displacement with length of 2.5 μm.

(**D**) Overlay of simulated displacement length distributions in 3D free space (PDF_{FS}, blue shade) and in confined space (PDF_{CS}, green shade). Apply ITCCD on PDF_{CS} recovers the displacement length distribution (red symbols) that agrees well with that in free space. Both the simulated PDF_{FS} and ITCDD match the theoretical displacement length distribution (black line) of the Brownian diffusion model. All distributions are normalized with the integrated area being one. (**E**, **F**), same as **B** and **D** but for 2D case. (**G**, **H**), same as **B** and **D** but the [CTM] is from 3D input displacement to 2D output displacement. ITCDD (red dots) clearly deviates from the theoretical displacement length distribution from the Brownian diffusion model (black line).

2.3.2 Analysis of one diffusion state in cells: variation in diffusion coefficient and cell geometry

The diffusion coefficient (D) of cytosolic molecules in bacteria typically ranges from 10^{-2} to 10 $\mu m^2 s^{-1}$.[88] To probe whether the magnitude of the diffusion coefficient may affect the performance of the inverse transformation method, we simulated diffusion trajectories with variable D's. As most of SMT experiments are done in 2D, we focus discussions and analyses on the 2D displacements generated from simulated diffusions that are always done in 3D; the results apply equally to the 3D displacements.

Figure 2A shows the simulated 2D PDF(r)'s in free space and in a confined cell volume with the input D of 11 $\mu m^2 s^{-1}$, a typical diffusion coefficient for a fast-diffusing small protein

in bacterial cytoplasm, for which the confinement effect is more significant than those with smaller diffusion coefficients. The corresponding ITCDD closely mimics that in free space (Figure A); fitting it with Eq (53) gives D_{fit} of 12.5 ± 0.2 $\mu m^2 s^{-1}$, within 14% of the input D. With the input D varying from 0.01 to 11 $\mu m^2 s^{-1}$, the fitted D from ITCDD is always within 0.1% to 14% of the input D (Figure B), smaller than or comparable to typical experimental uncertainties (8-25%)[60, 84]. Therefore, the inverse transformation method allows for direct and reliable extraction of the intrinsic diffusion coefficients of Brownian diffusions in confined space.

Using a fixed input D (e.g., 11 $\mu m^2 s^{-1}$), we further evaluated how the cell geometry, which the [CTM] is dependent on, might affect the performance of the inverse transformation method. We examined Brownian diffusions in cells with width of 1.15 μm and lengths of 2.32, 2.82, and 3.32 μm, corresponding to aspect ratios of 2.0, 2.5, and 2.9, respectively. These geometries cover the range of *E. coli* cell shapes typically observed in minimum growth medium[60]. Regardless of the cell geometry, the fitted D from ITCDD stays at 12.4 ± 0.5 $\mu m^2 s^{-1}$, within 13% of the input D (Figure C). To further test the insensitiveness of the inverse transformation method to cell geometry within this range, we combined the simulated distributions of displacement lengths from these three different geometries, but applied merely the [CTM] from the cells of length = 2.82 μm, which is about the average of the three cell lengths; fitting the ITCDD with Eq(53) again gives an D_{fit} of 12.3 ± 0.2 $\mu m^2 s^{-1}$, close to the input D. Therefore, even when diffusion trajectories are collected from a population of cells that differ in geometry (within the range evaluated here), it is sufficient to use the [CTM] for

the average cell geometry to perform ITCDD to extract the intrinsic diffusion coefficient.

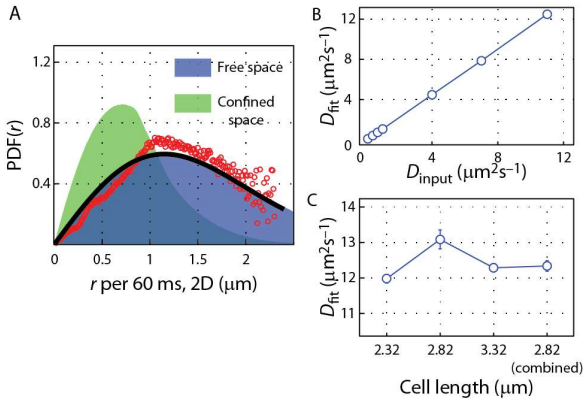


Figure 7. Analysis of one diffusion state inside a cell with ITCDD. (**A**) Simulated PDF(r) data of diffusion trajectories in free (blue shade) and confined (green shade) spaces with D_{input} of 11 μm^{2}s^{-1} and t = 60 ms. Multiplication of PDF(r) of diffusion trajectories in confined space with [CTM]$^{-1}$ resulted in the corresponding ITCDD (red circles), which reproduces the theoretical distribution (black curve) based on the Brownian model. (**B**) Fitted diffusion coefficients of ITCDD at various input diffusion coefficients. (**C**) Fitted diffusion coefficients of ITCDD (D_{input} of 11 μm^{2}s^{-1}) with cell lengths varying from 2.32 to 2.82 μm. The fit result of the combined ITCCD with the average cell length of 2.82 μm is also plotted for comparison.

2.3.3 Analysis of non-interconverting multi–state diffusions in cells

Inside a cell, a protein molecule may have a few different diffusive behaviors,

depending on its interactions with other proteins, RNA, or DNA. We thus evaluated the inverse transformation method in analyzing diffusion trajectories that contain multiple (i.e., three) Brownian diffusion states. We first examined the case that these states do not interconvert, i.e., a static mixture of diffusive behaviors. We again focus on the analysis of the 2D displacements here from 3D simulations. The diffusion coefficients (D_{input}) of diffusion states were set to 11 (D_1), 0.7 (D_2), and 0.036 (D_3) $\mu m^2 s^{-1}$, respectively, close to those we previously measured for a transcription factor in *E. coli* cells[60], and the fractional population (A_3) of the D_3 state was varied from 5 to 33% while the other two fractional populations were set as $A_1 = A_2 = (1-A_3)/2$.

Figure A shows the 2D PDF(r) from such a 3-state simulation in a cell volume. The corresponding ITCDD can be fitted using a linear combination of PDF(r) (Eq (55)), each accounting for one diffusion state with its corresponding fractional population (A) as a weighting coefficient:

$$\mathrm{PDF}_{3\mathrm{state}}(r,t) = \frac{A_1 r}{2D_1 t}\exp\left(-\frac{r^2}{4D_1 t}\right) + \frac{A_2 r}{2D_2 t}\exp\left(-\frac{r^2}{4D_2 t}\right) + \frac{(1-A_1-A_2)r}{2D_3 t}\exp\left(-\frac{r}{4l}\right) \quad (55)$$

The fitted D's of all three states are all within 10% error of D_{input} (Figure B). The fitted fractional populations are also in agreement with input ones (within 7% error), showing a clear trend of increasing A_3 and decreasing A_1 and A_2, as expected. Therefore, the inverse transformation method can effectively extract the intrinsic diffusion coefficients and their fractional populations of non-interconverting multi-state diffusions.

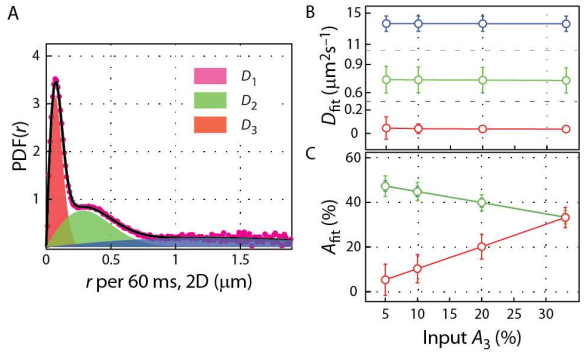


Figure 8. Analysis of non-interconverting multi-state diffusions inside a cell with ITCDD. (**A**) The ITCDD (magenta dots) from the simulated PDF(r) of diffusion trajectories in confined space (cell width and length of 1.15 and 2.82 μm) with three non-interconverting diffusion states with D_{input} of 11 (D_1), 0.7 (D_2), and 0.036 (D_3) $\mu m^2 s^{-1}$ and fractional populations of 33.3, 33.3, and 33.3%, respectively, and $t = 60$ ms. The overall fit with Eq (55) (black curve) and corresponding deconvoluted three diffusion states (blue, green, and red shades for D_1, D_2, and D_3 states, respectively) were overlaid. (**B**) Fitted diffusion coefficients of the three diffusion states from ITCDD (blue, green and red circles are for D_1, D_2, and D_3 states, respectively) when A_3 varies from 5 to 33%. Note the D_{fit} of each diffusion state was plotted in different y scales for clarity. (**C**) Fitted fractional populations of A_1 (blue circles), A_2 (green circles), and A_3 (red circles) from ITCDD at various A_3 inputs. Note the blue and green circles are on top of each other.

2.3.4 Analysis of interconverting multi–state diffusions in cells

Following the above section, we further evaluated the inverse transformation method in analyzing diffusion trajectories that contain three interconverting Brownian diffusion states. We simulated the 3D diffusion trajectories with 4 ms time resolution in a cell volume using a set of interconversion kinetic rate constants (Method Section 2.2) from our previous SMT study of the transcription factor CueR, which was tagged by a photoconvertible FP mEos3.2[89]. Table gives input parameters of this simulation, including diffusion coefficients (D_i, i = 1-3) and interconverting rate constants (γ_{ij}; i, j = 1, 2, or 3); the interconversion rate constants also determine the fractional populations of the respective states. We further included a photobleaching rate constant (k_{bl}) to account for that FP's photobleaching limits the length of tracking trajectories. Note that no interconversion was allowed between the D_2 and D_3 states, because it was kinetically negligible for CueR[60].

Focusing again on the analysis of 2D displacements, we first tested the inverse transformation method on the simulated trajectories at 60 ms time resolution (i.e., sample the displacement from the simulated diffusion trajectories with t = 60 ms). Figure A shows the ITCDD of the 2D PDF(r) in confined space from the simulation. Fitting it with Eq (55) gives D_1, D_2, and D_3 of 7.4 ± 3.1, 0.82 ± 0.09, and 0.037 ± 0.013 $\mu m^2 s^{-1}$; A_1, A_2, and A_3 of 29 ± 6, 55 ± 6, and 16 ± 8.5%, respectively. Compared with the simulation inputs, the ~10% error in the fitted diffusion coefficients and fractional populations again support that the inverse transformation method can effectively deconvolute the confinement effect.

One possible reason for the ~10% error in the fitted values of diffusion coefficient and fractional population of each state is the insufficient time resolution in sampling the simulated diffusion trajectories. t of 60 ms corresponds to a sampling rate of 15 s^{-1}, which is comparable to the interconversion rate constant γ_{13} between D_1 and D_3 state (Table). We therefore also analyzed the PDF(r) by sampling the simulated diffusion trajectories at t = 4 ms (Figure B). Fitting the ITCDD of this higher time resolution results with Eq (55) gives D_1, D_2, and D_3 of 12.1 ± 2.5, 0.74 ± 0.05, and 0.04 ± 0.01 $\mu m^2 s^{-1}$; A_1, A_2, and A_3 of 24 ± 3, 50 ± 2, and 26 ± 4% respectively, which now has ~3.6% error, significantly improved compared with those in analyzing the 60 ms time resolution results. Figure C compares the fractional populations from fitting ITCDD of 4 ms and 60 ms 2D displacement lengths with A_{inputs}. The results at 4 ms resolution can perfectly recover the correct fractional populations. Therefore, as long as the displacement is obtained at sufficient time resolution in SMT measurements, the inverse transformation of CDD is effective.

What time resolution (i.e. sampling rate) would then be sufficient? To address this, we systematically varied t from 4 to 60 ms in sampling the displacements in the simulated diffusion trajectories. Figure D shows the square-root of the sum of square error (i.e., SSE) of fractional populations ($SSE = \sum \Delta A^2$, $\Delta A = A_{\text{fit}} - A_{\text{input}}$) as a function of sampling time t. As t gets longer, $\sqrt{SSE}$ gets larger. Assuming $\sqrt{SSE}$ ($=\sqrt{\sum \Delta A^2}$) < 10% as being a good fit, t of 40 ms would be a minimum time resolution here for sampling the displacements so that the inverse transformation method would give the correct fractional populations of diffusion states. This t = 40 ms, corresponding to a rate of 25 s^{-1}, is about 2.2 times faster than the fastest

interconversion rate constant γ_{13} of 11.4 s^{-1} in the simulation.

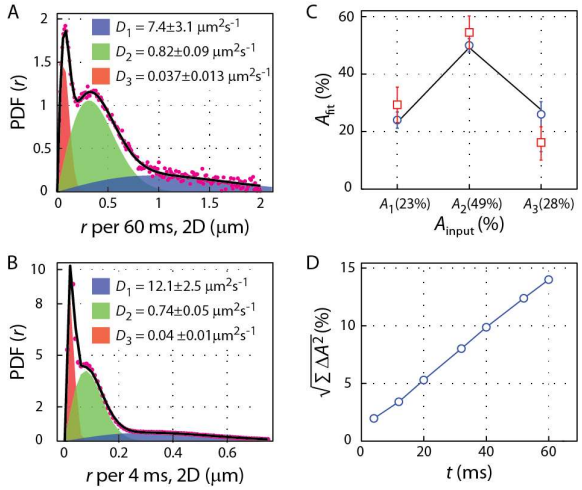
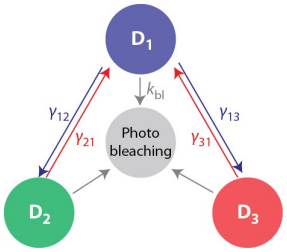


Figure 9. Analysis of interconverting multi-state diffusions inside a cell with ITCDD. (**A**) ITCDD (magenta dots) of the simulated PDF(r) of diffusion trajectories with 60 ms time resolution in confined space with three interconverting diffusion states with D_{input} in Table . The overall fit result (black curve) with Eq (55) and corresponding three diffusion states (blue, green, and red shades for D_1, D_2, and D_3 states, respectively) were overlaid. (**B**) Same as A, but with 4 ms time resolution. (**C**) Fitted fractional populations (i.e., A_1, A_2, and A_3) from ITCCD with 4 (blue circles) and 60 (red squares) ms time resolution, along with the input A_1, A_2, and A_3. (**D**) The $\sqrt{SSE}$ of fitted fractional populations from ITCCD at various time resolutions.

Table 2. Simulation Input parameters of three interconverting Brownian diffusion

states in a cell of 1.15 × 1.15 × 2.82 μm^3 in size

Parameters	Simulation inputs	Kinetic scheme
D_1 (μm^{2}s^{-1})	11	
D_2 (μm^{2}s^{-1})	0.7	
D_3 (μm^{2}s^{-1})	0.036	
γ_{12} (s^{-1})	5.3	
γ_{21} (s^{-1})	2.5	
γ_{13} (s^{-1})	11.4	
γ_{31} (s^{-1})	9.6	
k_{bl} (s^{-1})	252	
Theoretic A_1 (%)	23.2	
Theoretic A_2 (%)	49.2	
Theoretic A_3 (%)	27.6	

2.3.5 Application to transcription regulator dynamics in live E. coli cells

After validating the inverse transform method using simulated diffusion trajectories, we applied this method on the SMT data of CueR (in its apo form, i.e., apo-CueR), a Cu^+-responsive MerR-family transcription regulator, in living *E. coli* cells to extract the diffusion coefficient and fractional population of each diffusion state. Details on obtaining the SMT data were described in our previous work[60]. In short, we tagged the non-metallated apo-CueR with the photoconvertible FP mEos3.2, generating $\mathrm{CueR}^{\mathrm{mE}}_{\mathrm{apo}}$. We then used time-lapse stroboscopic

imaging to track the 2D motions of individual photoconverted $CueR_{apo}^{mE}$ in a cell at a sampling rate of every 60 ms until the mEos3.2 tag photobleached.

CueR can interact with DNA specifically at recognition sites or with DNA nonspecifically[90]. Three effective diffusion states are thus expected for $CueR_{apo}^{mE}$ in an *E. coli* cell: (1) specifically bound (SB) to chromosomal recognition sites, whose diffusion coefficient should be very small and largely reflect the chromosome conformational flexibility in the cell; (2) nonspecifically bound (NB) and moving along the chromosome; and (3) freely diffusing (FD) in the cytoplasm. FigureA shows the ITCDD from the measured 2D PDF(r) of $CueR_{apo}^{mE}$ at a low cellular protein concentration of ~100 nM (the cellular protein concentration was quantified for each cell in our imaging approach; details see our previous work[60]). Minimally three diffusion states are needed to fit the ITCDD satisfactorily. The three diffusion coefficients are 14.9 ± 7.6, 0.93 ± 0.08, and 0.062 ± 0.005 $\mu m^2 s^{-1}$, assignable as $CueR_{apo}^{mE}$ being FD in the cytoplasm, NB and SB to chromosome, respectively. On the other hand, at a high cellular protein concentration of ~1375 nM, the ITCDD only requires minimally two diffusion states to be fitted satisfactorily (FigureB). The two diffusion coefficients of 7.0 ± 1.5 and 0.90 ± 0.08 $\mu m^2 s^{-1}$ are within error to those of the FD and NB states at the low cellular protein concentration. Therefore, at this high cellular protein concentration, the SB state is no longer resolvable; this is not surprising since the fractional population of the SB state (i.e., (number of proteins specifically bound to DNA recognition sites/ (total number of proteins)) should be increasingly smaller at higher cellular protein concentrations. More important, these results demonstrate that depending on the cellular protein concentration, the experimentally resolvable

number of diffusion states can vary.

To more reliably determine the minimal number of diffusion states, we propose using a global fit of ITCDD across all cellular protein concentrations (i.e., four sets of data, each set coming from a sorted group of cells with a particular cellular protein concentration[60]), where the number of diffusion states and their diffusion coefficients are shared. This global fit on the results of $\mathrm{CueR}^{\mathrm{mE}}_{\mathrm{apo}}$ gave three states with diffusion coefficients of 8.2 ± 0.3, 0.92 ± 0.04, and 0.051 ± 0.005 $\mu m^2 s^{-1}$, corresponding to the FD, NB, and SB states, respectively. Compared with literature values[61, 74, 84, 91], they are in excellent agreements with those expected for freely diffusing in the cytoplasm, nonspecifically bound to chromosome, and specifically bound to chromosome. FigureC summarizes the extracted fractional populations as a function of cellular protein concentration. With increasing protein concentration, the fractional population of the SB state decreases while those of NB and FD states increase, consistent with expectations and our previous study[60], and further supporting the effectiveness of globally fitting the ITCDD.

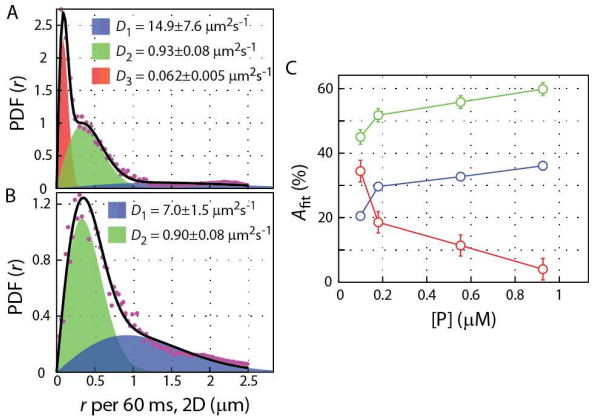


Figure 10. Analysis of diffusive behaviors of apo-CueR in *E. coli* cells with ITCDD. (**A**) ITCDD (magenta dots) from experimental PDF(*r*) data of apo-CueR with 60 ms sampling rate in cells with total cellular protein concentration ([apo-CueR]) of 99 nM. The ITCDD requires a model with three diffusion states to achieve satisfactory fitting. The overall fit result (black curve) and corresponding three resolved diffusion states (blue, green, and red shades for D_1, D_2, and D_3 state respectively) were overlaid. (**B**) Same as A, but with total [apo-CueR] = 1375 nM. Two diffusion states are sufficient to fit the ITCDD satisfactorily. (**C**) Fitted fractional populations of A_1, A_2, and A_3 of ITCDD for [apo-CueR] varying from 99 to 926 nM, in which a global fit across 4 sets of data of different cellular protein concentrations was performed. The [CTM] was based on the experimentally determined cell geometry of cell width and length of 1.15 and 2.82 μm.

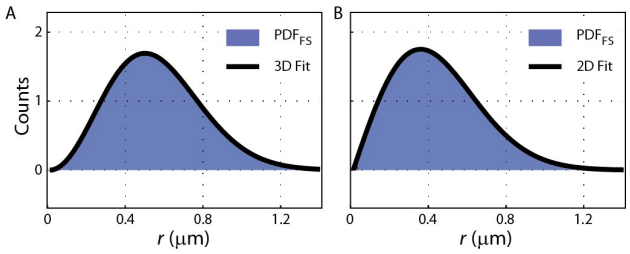
2.4 Conclusion

High spatial and temporal resolution position trajectories from SMT of fluorescently-tagged cytoplasmic proteins carry valuable information about the underlying biological

processes in cells but their analysis is complicated by the confinement effect from the cell volume, especially for small bacterial cells. Here we deconvolute out the confinement effect by inverse transforming the probability density function of displacement length (PDF(r)) using the confinement transformation matrix ([CTM]), building on the earlier work of Peterman[87] on simulated single-state membrane diffusions. Besides treating single-state cytoplasmic diffusions, we further extended this method to analyze multi-state Brownian diffusions in the cytoplasm, including both non-interconverting and interconverting three-state diffusions. We demonstrated the effectiveness of this method in determining the minimal number of diffusion states, their diffusion coefficients and fractional populations, as well as how to choose a sufficient time resolution in analyzing systems containing inter-converting multi-states. A successful application to experimental multi-state SMT data of a transcription factor in live *E. coli* cell is also demonstrated. Together with Peterman's early work on membrane diffusion (whose extension to multi-state systems can readily follow our work on cytoplasmic diffusion here), our method allows for direct connection between SMT data with diffusion theory for analyzing molecular diffusive behaviors in live bacteria.

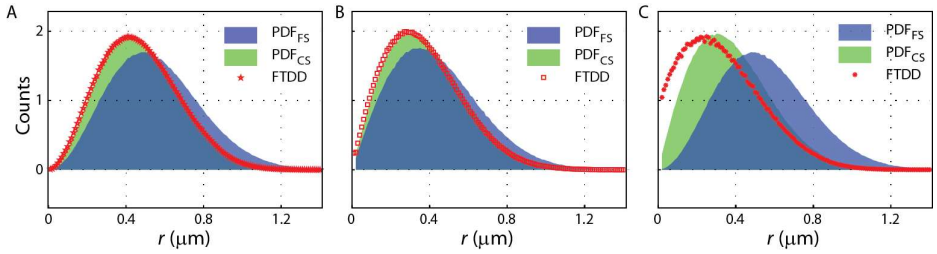
2.5 Supporting information to chapter 2

2.5.1 Validation of Brownian diffusion simulation results



Supplementary Figure 19. Displacement length distributions and fittings from simulated 3D and 2D diffusion data. (**A**) Displacement length distribution (PDF_{FS}) constructed from the 3D diffusion simulation data in free space. The distribution can be well fitted by Eq (52) (black curve), giving the diffusion coefficient of 1.001 ± 0.001 $\mu m^2 s^{-1}$, identical to the D_{input} of 1 $\mu m^2 s^{-1}$. (B) Same as A, but from 2D diffusion simulation data. The distribution can also be well fitted by Eq (53) (black curve), giving the diffusion coefficient of 1.001 ± 0.001, i.e., identical to the expected value of 1 $\mu m^2 s^{-1}$.

2.5.2 Forward transformation of displacement distribution (FTDD) in free space with [CTM]



Supplementary Figure 20. Forward transformation of displacement length distribution in free space into that in confined space using [CTM]. (**A**) The corresponding

forward transformation results of Figure D. The simulated displacement distributions in free space (blue shade), confined space (green shade), and forward transformation result (FTDD, red symbol) were plotted. (**B**) Same as A, but corresponds to Figure F. (**C**) Same as A, but corresponds to Figure H. Here the forward transformation also shows a clear deviation from displacement distribution in confined space because of dimension mismatch.

REFERENCES

1. Schramm, L. & Hernandez, N. Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16**, 2593-2620 (2002).
2. Deuschle, U., Gentz, R. & Bujard, H. lac Repressor blocks transcribing RNA polymerase and terminates transcription. *Proc. Natl Acad. Sci. USA* **83**, 4134-4137 (1986).
3. Rajagopalan, S. et al. Studies of IscR reveal a unique mechanism for metal-dependent regulation of DNA binding specificity. *Nat. Struct. Mol. Biol.* **20**, 740-747 (2013).
4. Nesbit, A. D., Giel, J. L., Rose, J. C. & Kiley, P. J. Sequence-specific binding to a subset of IscR-regulated promoters does not require IscR Fe-S cluster ligation. *J. Mol. Biol.* **387**, 28-41 (2009).
5. Chen, T. Y., Cheng, Y. S., Huang, P. S. & Chen, P. Facilitated Unbinding via Multivalency-Enabled Ternary Complexes: New Paradigm for Protein-DNA Interactions. *Acc. Chem. Res.* **51**, 860-868 (2018).
6. Graham, J. S., Johnson, R. C. & Marko, J. F. Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA. *Nucleic Acids Res.* **39**, 2249-2259 (2011).
7. Gibb, B. et al. Concentration-dependent exchange of replication protein A on single-stranded DNA revealed by single-molecule imaging. *PLoS One* **9**, e87922 (2014).
8. Geertsema, H. J., Kulczyk, A. W., Richardson, C. C. & van Oijen, A. M. Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery. *Proc Natl Acad Sci U S A* **111**, 4073-4078 (2014).
9. Lewis, J. S. et al. Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *Elife* **6** (2017).
10. Joshi, C. P. et al. Direct substitution and assisted dissociation pathways for turning off transcription by a MerR-family metalloregulator. *Proc. Natl Acad. Sci. USA* **109**, 15121-15126 (2012).
11. Chen, T. Y. et al. Concentration- and chromosome-organization-dependent regulator unbinding from DNA for transcription regulation in living cells. *Nat. Commun.* **6**, 7445 (2015).
12. Hantke, K. Bacterial zinc uptake and regulators. *Curr. Opin. Microbiol.* **8**, 196-202 (2005).
13. Hemm, M. R. et al. Small stress response proteins in Escherichia coli: proteins missed by classical proteomic studies. *J. Bacteriol.* **192**, 46-58 (2010).
14. Panina, E. M., Mironov, A. A. & Gelfand, M. S. Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proc. Natl Acad. Sci. USA* **100**, 9912-9917 (2003).
15. Gilston, B. A. et al. Structural and mechanistic basis of zinc regulation across the E. coli Zur regulon. *PLoS Biol.* **12**, e1001987 (2014).
16. Zhang, M. et al. Rational design of true monomeric and bright photoactivatable fluorescent proteins. *Nat. Methods* **9**, 727-729 (2012).
17. McKinney, S. A. et al. A bright and photostable photoconvertible fluorescent protein. *Nat. Methods* **6**, 131-133 (2009).
18. Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-

molecule level in a living cell. *Science* **316**, 1191-1194 (2007).
19. Javer, A. et al. Short-time movement of E. coli chromosomal loci depends on coordinate and subcellular localization. *Nat. Commun.* **4**, 3003 (2013).
20. Mehta, P. et al. Dynamics and stoichiometry of a regulated enhancer-binding protein in live Escherichia coli cells. *Nat. Commun.* **4**, 1997 (2013).
21. Uphoff, S. et al. Single-molecule DNA repair in live bacteria. *Proc. Natl Acad. Sci. USA* **110**, 8063-8068 (2013).
22. Mazza, D., Ganguly, S. & McNally, J. G. Monitoring dynamic binding of chromatin proteins in vivo by single-molecule tracking. *Methods Mol. Biol.* **1042**, 117-137 (2013).
23. Ma, Z., Gabriel, S. E. & Helmann, J. D. Sequential binding and sensing of Zn(II) by Bacillus subtilis Zur. *Nucleic Acids Res.* **39**, 9130-9138 (2011).
24. Chen, P. et al. Single-molecule dynamics and mechanisms of metalloregulators and metallochaperones. *Biochemistry* **52**, 7170-7183 (2013).
25. Choi, S. H. et al. Zinc-dependent regulation of zinc import and export genes by Zur. *Nat. Commun.* **8**, 15812 (2017).
26. Wang, W. et al. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333**, 1445-1449 (2011).
27. Prestel, E., Noirot, P. & Auger, S. Genome-wide identification of Bacillus subtilis Zur-binding sites associated with a Zur box expands its known regulatory network. *BMC Microbiol.* **15**, 13 (2015).
28. Hadizadeh, N., Johnson, R. C. & Marko, J. F. Facilitated Dissociation of a Nucleoid Protein from the Bacterial Chromosome. *J. Bacteriol.* **198**, 1735-1742 (2016).
29. Geertsema, H. J., Kulczyk, A. W., Richardson, C. C. & van Oijen, A. M. Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery. *Proc. Natl Acad. Sci. USA* **111**, 4073-4078 (2014).
30. Norel, R., Sheinerman, F., Petrey, D. & Honig, B. Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci.* **10**, 2147-2161 (2001).
31. Xu, D., Tsai, C. J. & Nussinov, R. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.* **10**, 999-1012 (1997).
32. Zhang, Z., Witham, S. & Alexov, E. On the role of electrostatics in protein-protein interactions. *Phys. Biol.* **8**, 035001 (2011).
33. Persson, B. A. & Lund, M. Association and electrostatic steering of alpha-lactalbumin-lysozyme heterodimers. *Phys. Chem. Chem. Phys.* **11**, 8879-8885 (2009).
34. Gunasekaran, K. et al. Enhancing antibody Fc heterodimer formation through electrostatic steering effects: applications to bispecific molecules and monovalent IgG. *J. Biol. Chem.* **285**, 19637-19646 (2010).
35. Persson, B. A., Jonsson, B. & Lund, M. Enhanced protein steering: cooperative electrostatic and van der Waals forces in antigen-antibody complexes. *J. Phys. Chem. B* **113**, 10459-10464 (2009).
36. Hemsath, L. et al. An electrostatic steering mechanism of Cdc42 recognition by Wiskott-Aldrich syndrome proteins. *Mol. Cell* **20**, 313-324 (2005).
37. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640-6645 (2000).
38. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation,

and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121-4130 (1995).
39. Thompson, R. E., Larson, D. R. & Webb, W. W. Precise Nanometer Localization Analysis for Individual Fluorescent Probes. *Biophys. J.* **82**, 2775-2783 (2002).
40. Durisic, N. et al. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. *Nat. Methods* **11**, 156-162 (2014).
41. Fick, A. V. On liquid diffusion. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **10**, 30-39 (1855).
42. Gebhardt, J. C. et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods* **10**, 421-426 (2013).
43. English, B. P. et al. Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proc. Natl Acad. Sci. USA* **108**, E365-373 (2011).
44. Mazza, D. et al. A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Res.* **40**, e119 (2012).
45. Niu, L. & Yu, J. Investigating intracellular dynamics of FtsZ cytoskeleton with photoactivation single-molecule tracking. *Biophys. J.* **95**, 2009-2016 (2008).
46. Mueller, F., Stasevich, T. J., Mazza, D. & McNally, J. G. Quantifying transcription factor kinetics: at work or at play? *Crit. Rev. Biochem. Mol. Biol.* **48**, 492-514 (2013).
47. Chen, T. Y. et al. Quantifying Multistate Cytoplasmic Molecular Diffusion in Bacterial Cells via Inverse Transform of Confined Displacement Distribution. *J. Phys. Chem. B* **119**, 14451-14459 (2015).
48. Brunauer, S., Emmett, P. H. & Teller, E. Adsorption of Gases in Multimolecular Layers. *J. Am. Chem. Soc.* **60**, 309-319 (1938).
49. Gao, W., Zhang, W. & Meldrum, D. R. RT-qPCR based quantitative analysis of gene expression in single bacterial cells. *J. Microbiol. Methods* **85**, 221-227 (2011).
50. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nature protocols* **3**, 1101-1108 (2008).
51. Bakshi, S., Bratton, B. P. & Weisshaar, J. C. Subdiffraction-limit study of Kaede diffusion and spatial distribution in live Escherichia coli. *Biophys. J.* **101**, 2535-2544 (2011).
52. Annibale, P., Scarselli, M., Greco, M. & Radenovic, A. Identification of the factors affecting co-localization precision for quantitative multicolor localization microscopy. *Optical Nanoscopy* **1** (2012).
53. Yildiz, A. et al. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* **300**, 2061-2065 (2003).
54. Cunha, S., Woldringh, C. L. & Odijk, T. Restricted diffusion of DNA segments within the isolated Escherichia coli nucleoid. *Journal of structural biology* **150**, 226-232 (2005).
55. Golding, I. & Cox, E. C. Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.* **96**, 098102 (2006).
56. Yang, F. et al. Single-molecule dynamics of the molecular chaperone trigger factor in living cells. *Mol. Microbiol.* **102**, 992-1003 (2016).
57. Andoy, N. M. et al. Single-molecule study of metalloregulator CueR-DNA interactions using engineered Holliday junctions. *Biophys. J.* **97**, 844-852 (2009).
58. Leake, M. C. et al. Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging. *Proceedings of*

the National Academy of Sciences of the United States of America **105**, 15376-15381 (2008).
59. Lillemeier, B. F. et al. Plasma membrane-associated proteins are clustered into islands attached to the cytoskeleton. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18992-18997 (2006).
60. Chen, T.-Y. et al. Concentration- and chromosome-organization-dependent regulator unbinding from DNA for transcription regulation in living cells. *Nat. Commun.* **6**, DOI: 10.1038/ncomms8445 (2015).
61. Uphoff, S. et al. Single-molecule DNA repair in live bacteria. *Proceedings of the National Academy of Sciences* **110**, 8063-8068 (2013).
62. Niu, L. & Yu, J. Investigating Intracellular Dynamics of FtsZ Cytoskeleton with Photoactivation Single-Molecule Tracking. *Biophysical Journal* **95**, 2009-2016 (2008).
63. Javer, A. et al. Short-time movement of E. coli chromosomal loci depends on coordinate and subcellular localization. *Nat. Commun.* **4**, DOI: 10.1038/ncomms3003 (2013).
64. Axelrod, D. et al. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophysical Journal* **16**, 1055-1069 (1976).
65. Hess, S. T., Huang, S., Heikal, A. A. & Webb, W. W. Biological and Chemical Applications of Fluorescence Correlation Spectroscopy: A Review. *Biochemistry* **41**, 697-705 (2002).
66. Patterson, G., Davidson, M., Manley, S. & Lippincott-Schwartz, J. Superresolution Imaging using Single-Molecule Localization. *Annual Review of Physical Chemistry* **61**, 345-367 (2010).
67. Mueller, F., Stasevich, T. J., Mazza, D. & McNally, J. G. Quantifying transcription factor kinetics: at work or at play? *Crit Rev Biochem Mol Biol* **48**, 492-514 (2013).
68. Solarczyk, K. J., Zarębski, M. & Dobrucki, J. W. Inducing local DNA damage by visible light to study chromatin repair. *DNA Repair* **11**, 996-1002 (2012).
69. Ries, J. & Schwille, P. Fluorescence correlation spectroscopy. *Bioessays* **34**, 361-368 (2012).
70. Small, A. & Stahlheber, S. Fluorophore localization algorithms for super-resolution microscopy. *Nat. Meth.* **11**, 267-279 (2014).
71. Hiramoto-Yamaki, N. et al. Ultrafast Diffusion of a Fluorescent Cholesterol Analog in Compartmentalized Plasma Membranes. *Traffic* **15**, 583-612 (2014).
72. Fernandez-Suarez, M. & Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nat. Rev. Mol. Cell. Biol.* **9**, 929-943 (2008).
73. Yu, J. et al. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. *Science* **311**, 1600-1603 (2006).
74. Elf, J., Li, G.-W. & Xie, X. S. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **316**, 1191-1194 (2007).
75. Pinaud, F., Clarke, S., Sittner, A. & Dahan, M. Probing cellular events, one quantum dot at a time. *Nat. Meth.* **7**, 275-285 (2010).
76. Beausang, J. F. et al. DNA Looping Kinetics Analyzed Using Diffusive Hidden Markov Model. *Biophysical Journal* **92**, L64-L66 (2007).
77. Chung, I. et al. Spatial control of EGF receptor activation by reversible dimerization on living cells. *Nature* **464**, 783-787 (2010).
78. Das, R., Cairo, C. W. & Coombs, D. A Hidden Markov Model for Single Particle Tracks Quantifies Dynamic Interactions between LFA-1 and the Actin Cytoskeleton. *PLoS*

Comput. Biol. **5**, DOI: 10.1371/journal.pcbi.1000556 (2009).
79. Persson, F., Linden, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat. Meth.* **10**, 265-269 (2013).
80. Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Physical Review E* **82**, 041914 (2010).
81. Bosch, Peter J., Kanger, Johannes S. & Subramaniam, V. Classification of Dynamical Diffusion States in Single Molecule Tracking Microscopy. *Biophysical Journal* **107**, 588-598 (2014).
82. Robson, A., Burrage, K. & Leake, M. C. Inferring diffusion in single live cells at the single-molecule level. *Philosophical Transactions B* **368** (2013).
83. Blanco, M., Johnson-Buck, A. & Walter, N. Hidden Markov Modeling in Single-Molecule Biophysics, in *Encyclopedia of Biophysics*. (ed. G. K. Roberts) 971-975 (Springer Berlin Heidelberg, 2013).
84. Gebhardt, J. C. M. et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Meth.* **10**, 421-426 (2013).
85. Morisaki, T. et al. Single-molecule analysis of transcription factor binding at transcription sites in live cells. *Nat. Commun.* **5**, DOI: 10.1038/ncomms5456 (2014).
86. Bakshi, S., Choi, H., Mondal, J. & Weisshaar, J. C. Time-dependent effects of transcription- and translation-halting drugs on the spatial distributions of the Escherichia coli chromosome and ribosomes. *Molecular Microbiology* **94**, 871-887 (2014).
87. Oswald, F., L. M. Bank, E., Bollen, Y. J. M. & Peterman, E. J. G. Imaging and quantification of trans-membrane protein diffusion in living bacteria. *Physical Chemistry Chemical Physics* **16**, 12625-12634 (2014).
88. Elowitz, M. B. et al. Protein Mobility in the Cytoplasm ofEscherichia coli. *Journal of Bacteriology* **181**, 197-203 (1999).
89. Zhang, M. et al. Rational design of true monomeric and bright photoactivatable fluorescent proteins. *Nat. Meth.* **9**, 727-729 (2012).
90. Joshi, C. P. et al. Direct Substitution and Assisted Dissociation Pathways for Turning Off Transcription by A MerR-family Metalloregulator. *Proc. Nat. Acad. Sci. USA* **109**, 15121-15126 (2012).
91. Mehta, P. et al. Dynamics and stoichiometry of a regulated enhancer-binding protein in live Escherichia coli cells. *Nat. Commun.* **4**, DOI: 10.1038/ncomms2997 (2013).